

Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)

Stereorekonstruktion mit Anwendung auf die Konstruktion eines Verbundes von Stereokameras

P. Schmid

Stereorekonstruktion mit Anwendung auf die Konstruktion eines Verbundes von Stereokameras

P. Schmid

Berichte des Forschungszentrums Jülich; 4341
ISSN 0944-2952
Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)
Jül-4341

Vollständig frei verfügbar im Internet auf dem Jülicher Open Access Server (JUWEL)
unter <http://www.fz-juelich.de/zb/juwel>

Zu beziehen durch: Forschungszentrum Jülich GmbH · Zentralbibliothek, Verlag
D-52425 Jülich · Bundesrepublik Deutschland
☎ 02461 61-5220 · Telefax: 02461 61-6103 · e-mail: zb-publikation@fz-juelich.de

Abstract

At the Jülich Supercomputing Centre at the Research Centre Jülich are not only the operations of the supercomputers but also mathematical modeling and scientific simulation a large working area. A special project is to model movements of pedestrians. Inter alia movements and routes of people are recorded by single or stereo cameras in order to get images and data that permit developing and verifying an appropriate model.

The disadvantage of conventional and commercial stereo cameras is the fixed alignment of the two lenses to each other so they can't individually be adjusted to each other. But in many cases a higher variability and individual adjustment of the lenses are requested, so constructing a stereo cameras by two independant cameras is more advantageously. This stereo system will be flexible in adjustment, for example concerning the distance between the two lenses.

The two monocular cameras have to be calibrated and aligned correctly to each other. Methods of the digital image processing and camera technology are necessary for calibrating a stereo system: Undistortion, Rectification etc.. Afterwards with the calibrated stereo camera, stereo images can be recorded in order to get information about the distance between a point and the camera. Calculating the distance takes over disparities which describe the difference between the coordinates of two corresponding image points. Then the distance from a pixel to the camera can mathematically defined

as proportional inverse to disparity.

Besides the construction of an individual stereo camera, the result of this paper is the generation of a network of stereo cameras, called stereo grid. This assures an enhanced observing area which exceeds one camera visual range and allows to observe larger scenes. The experiments has shown that the construcion af a stereo grid can be realized by using the previously acquired knowledge and the models of stereo calibration. An accurate adjustment of the cameras can be achieved due to the stereo calibration and the undistortion which provides the intrinsic and extrinsic parameters. Compared to a single stereo system, the geometrical relationships and dependencies of the cameras to each other have to be extended and adjusted, but they are still determined through the explained perspective transformation.

Abstrakt

Im Jülich Supercomputing Centre des Forschungszentrum Jülichs ist neben dem Betrieb von Superrechnern, die mathematische Modellierung und wissenschaftliche Simulation ein großes Arbeitsgebiet. Ein Projekt ist dabei die Modellierung der Bewegungen von Fußgängern. Unter anderem werden dabei Laufwege von einzelnen Personen durch Einzel- oder Stereokameras erfasst, welche Aufnahmen und Daten liefern, die eine geeignete Modellentwicklung und Modellverifikation ermöglichen.

Bei den herkömmlichen, im Handel erhältlichen Stereokameras ist das Problem, dass die beiden Linsensysteme fest zueinander ausgerichtet sind. Da aber in vielen Fällen eine höhere Variabilität und eine individuelle Einstellung gefordert ist, soll ein Stereosystem aus zwei unabhängigen Kameras konstruiert werden. Dieses ist dann flexibel, zum Beispiel im Abstand der zwei Linsen zueinander, einstellbar. Die zwei Kameras müssen kalibriert und entsprechend zueinander ausgerichtet werden. Nötig sind dabei Verfahren aus der digitalen Bildverarbeitung undameratechnik: Entzerrung, Rektifizierung etc.. Mit Hilfe des kalibrierten Stereosystems können anschließend Stereoaufnahmen gemacht werden, die es zulassen, Informationen über den Abstand eines Punktes zur Kamera zu gewinnen. Die Berechnung des Abstandes geschieht über Disparitäten, welche die Differenz zwischen den Koordinaten zweier korrespondierender Bildpunkte beschreibt. Der Abstand eines Pixel zur Kamera lässt sich dann mathematisch als

invers proportional zur Disparität definieren.

Das Ergebnis dieser Ausarbeitung ist, neben der Konstruktion eines individuellen Stereosystems, die Erzeugung eines Verbundes aus Kameras, einem Stereogrid. Dieses gewährleistet eine vergrößerte Beobachtungsfläche, die über einen Kamerasichtbereich hinausgeht und eine Observierung von größere Szenen ermöglicht.

Die Arbeit hat gezeigt, dass mittels der vorher erlangten Erkenntnisse und Modelle zur Stereokalibrierung die Konstruktion des Stereogrids verständlich und zu realisieren sind. Durch die vorgestellte Stereokalibrierung, die im Wesentlichen durch die Entzerrung und der damit verbundenen Berechnung der intrinsischen und extrinsischen Parameter geschieht, kann auch die korrekte Ausrichtung der Kameras im Stereogrid erreicht werden. Die geometrischen Beziehungen und die Abhängigkeiten der Kameras zueinander müssen im Vergleich zum einzelnen Stereosystem erweitert und angepasst werden, sind aber dennoch über die erläuterte perspektivische Transformation bestimmbar.

Inhaltsverzeichnis

| | | |
|----------|--------------------------------------|-----------|
| 1 | Einleitung | 1 |
| 1.1 | Motivation | 1 |
| 1.2 | Einführung | 2 |
| 2 | Stereokalibrierung | 5 |
| 2.1 | Lochkameramodell | 9 |
| 2.2 | Verzerrungen | 15 |
| 2.2.1 | Radiale Verzerrung | 15 |
| 2.2.2 | Tangentiale Verzerrung | 17 |
| 2.3 | Entzerrung | 18 |
| 2.4 | Epipolargeometrie | 23 |
| 2.5 | Rektifizierung | 29 |
| 3 | Disparitäten | 33 |
| 3.1 | Block Matching | 38 |
| 3.2 | Semi-Global Block Matching | 42 |

| | | |
|----------|-------------------------------------|-----------|
| 4 | Verbund von Stereokameras | 47 |
| 4.1 | Aufbau | 48 |
| 4.2 | Konstruktion | 49 |
| 5 | Zusammenfassung und Ausblick | 53 |
| 5.1 | Ausblick | 57 |
| A | Implementierung | 59 |
| B | Tabellen | 63 |

Abbildungsverzeichnis

| | | |
|------|---|----|
| 2.1 | Schachbrett als Kalibrierungsobjekt | 7 |
| 2.2 | Die Lochkamera | 9 |
| 2.3 | Das Lochkameramodell | 10 |
| 2.4 | Perspektivische Transformation des Welt- ins Kamerako- ordinatensystem | 10 |
| 2.5 | Bestimmung der internen Eckpunkte | 13 |
| 2.6 | Beispiel für radiale Verzerrung | 15 |
| 2.7 | Prinzip der radialen Verzerrung | 16 |
| 2.8 | Beispiel für die tangentielle Verzerrung | 17 |
| 2.9 | Prinzip der tangentialen Verzerrung | 18 |
| 2.10 | Verzerrtes Bild | 20 |
| 2.11 | Entzerrtes Bild | 21 |
| 2.12 | Epipolargeometrie | 24 |
| 2.13 | Koordinatentransformation der Epipolargeometrie | 26 |
| 2.14 | Rektifiziertes Stereokamerasystem | 29 |
| 2.15 | Vor der Rektifizierung | 31 |

| | | |
|------|---|----|
| 2.16 | Nach der Rektifizierung | 31 |
| 3.1 | Korrespondenzanalyse | 33 |
| 3.2 | Zusammenhang Disparität und Abstand zur Kamera | 34 |
| 3.3 | Disparität und Abstand | 36 |
| 3.4 | Problematik bei der Disparitätsbestimmung | 38 |
| 3.5 | Funktionsweise des Block Matching Algorithmus | 40 |
| 3.6 | Rektifiziertes Stereobildpaar | 41 |
| 3.7 | Disparitätsbild durch BM-Algorithmus | 41 |
| 3.8 | Pfadanalyse | 44 |
| 3.9 | Disparitätsbild durch den SGBM-Algorithmus | 45 |
| 4.1 | Kameragridausrichtung | 48 |
| 4.2 | Kameraanbringung | 49 |
| 4.3 | Bilden von Kamerapaaren | 51 |
| 4.4 | Gridaufnahme | 52 |
| A.1 | Stereorekonstruktion | 60 |
| A.2 | Stereogridkonstruktion | 62 |

Kapitel 1

Einleitung

1.1 Motivation

Motiviert wird die Thematik unter anderem durch den im Jülich Supercomputing Centre des Forschungszentrums Jülichs entwickelten Evakuierungsassistenten, der die Sicherheit bei Großveranstaltungen verbessern soll. Er simuliert das Verhalten von Menschen im Evakuierungsnotfall im Voraus, sodass frühzeitig zum Beispiel Staus prognostiziert und Ordnungskräfte gezielt eingesetzt werden können. Um die Simulation zu verifizieren, werden Daten benötigt, welche mit Hilfe von mehreren Kameras bei Experimenten erfasst werden. Bei den bisherigen Experimenten wurden die mit den Kameras verfolgten Personen mit einer speziellen Haube markiert, um ihre Position zu verfolgen.

Personen sollen nun aber auch ohne Markierung erkannt werden. Dies soll geschehen, indem durch aufgenommene Stereobilder Höhenbilder berechnet werden, die Informationen über den Abstand jeden Pixels zur Kamera liefern. Durch die Analyse der Höhenbilder ist es möglich, Personen zu erkennen und diese von Wänden, Gängen oder Treppen zu unterscheiden. In dieser Arbeit wird der nötige Schritt der Stereorekonstruktion implementiert und ausgearbeitet. Eine individuelle Stereokamera wird aus zwei monokularen Kameras konstruiert. Der Vorteil eines selbst erstellten Ste-

reosystems ist es, dass die beiden Linsensysteme individuell einstellbar und variabel in Ausrichtung und Position sind. Sie sind nicht fest eingebaut und können auf persönliche Anforderungen abgestimmt werden. Zur Konstruktion der Stereokamera ist eine korrekte Ausrichtung zweier monolukarer Kameras notwendig. Sie müssen so ausgerichtet werden, dass sie zwar verschiedene Blickwinkel zur Szene haben, dennoch muss sich ihr Sichtbereich größtenteils überschneiden. Ein Verbund von Stereokameras, im Folgenden als Stereogrid bezeichnet, soll eine Beobachtungsfläche vergrößern. Dies ist eine Vorarbeit zur Realisierung der Verfolgung von Personen über einen Kamerasichtbereich hinaus.

Ziel dieser Arbeit ist es somit, zum Einen aus zwei monokularen Kameras ein Stereosystem zu konstruieren. Die beiden Kameras müssen entsprechend ausgerichtet und relevante Parameter angepasst werden, um anschließend Disparitätsinformationen aus dem Stereosystem ermitteln zu können. Zum Anderen soll aus Kameras ein Grid erstellt werden, das eine vergrößerte Beobachtungsfläche der Szene zulässt.

1.2 Einführung

Zur Erarbeitung der Thematik und Implementierung des Modells stehen drei Kameras zur Verfügung. Zwei Kameras werden dabei zu einer Stereokamera zusammengefasst, welche wie das menschliche Auge „sieht“, so dass auch dreidimensionales Sehen möglich ist. Zur entsprechenden Weiterverarbeitung der aufgenommenen Bilder sind Verfahren und Techniken aus der Bildverarbeitung undameratechnik notwendig: Entzerrung, Rektifizierung, Korrespondenzanalyse, perspektivische Transformation, Kamerakalibrierung etc..

Die dritte Kamera wird zusätzlich für die Konstruktion eines elementaren Kameragrids benötigt. Die drei Kameras müssen entsprechend kalibriert und ausgerichtet werden, um eine vergrößerte Beobachtungsfläche zu bieten.

Die verfügbaren Kameras des Herstellers IDS GmbH können über USB 2.0 mit einem Computer verbunden werden. Zwei dieser Kameras gehö-

ren zum gleichen Modell *UI – 2230SE*. Dieses Modell ist eine Kamera mit einem CCD-Sensor in XGA-Auflösung von Sony [17]. Diese beiden Kameras werden aufgrund der Gleichheit für die Kalibrierung des Stereosystems verwendet.

Die dritte Kamera ist ebenfalls vom Hersteller IDS GmbH und aus der gleichen Serie, dennoch gehört sie zu dem neueren Modell *UI – 1240SE*. Sie besitzt anstatt eines CCD-Sensors einen CMOS-Sensor mit einer höheren Auflösung [17], sodass sich die beiden vorgestellten Kameramodelle im Wesentlichen durch ihr Objektiv unterscheiden. Weitere Eigenschaften werden in der Tabelle B.1 aufgelistet.

Die Implementierung der Arbeit erfolgt in C++ mit Hilfe der Bibliothek QT Version 4.7.2 zur plattformübergreifenden Programmierung der graphischen Oberfläche [6]. Desweiteren wurde die Opensource-Bibliothek OpenCV Version 2.2 verwendet, die für das Bildverstehen (Computer Vision) entwickelt wurde und schon viele Algorithmen und Funktionen zur Stereokalibrierung sowie Bildverarbeitung zur Verfügung stellt [2]. Weitere Details zur Implementierung sind im Anhang A zu finden.

Im ersten Kapitel *Stereokalibrierung* wird geschildert, welche Verfahrensschritte zur Korrektur von Bildern, zur Vereinfachung von geometrischen Beziehungen und zur Bestimmung von wichtigen Parametern für die Stereokalibrierung notwendig sind. Dabei werden relevante Modelle vorgestellt und wichtige Zusammenhänge verdeutlicht. Im anschließenden Kapitel *Disparitätsermittlung* wird aus Stereobildern eine Disparitätsmatrix bestimmt, aus der Tiefeninformationen gewonnen werden können. Dafür werden zwei in OpenCV implementierte Algorithmen erklärt und kurz miteinander verglichen. Der Bezug von Disparität zum Abstand eines Pixels zur Kamera wird ebenfalls aufgezeigt. Im nächsten Kapitel, welches die Konstruktion des Stereogrids behandelt, werden die in den vorherigen Kapiteln vorgestellten Schritte und Verfahren miteinbezogen und kurz zusammengefasst. An ausgewiesenen Stellen werden die bisherigen Modelle angepasst und erweitert. Zum Schluss gibt es neben einer Zusammenfassung

der erläuterten Vorgehensweisen zur Konstruktion eines Stereogrids einen Ausblick, der zu weiterführenden Arbeiten und Experimenten anregen soll.

Kapitel 2

Stereokalibrierung

Die Herausforderung bei einem Stereosystem ist es, korrespondierende Punkte in verschiedenen Perspektiven genau zu identifizieren, da sie in den beiden entstehenden Bildern an unterschiedlichen Stellen liegen können und die aufgenommenen Bilder durch bestimmte Parameter verzerrt sind. Daher muss die Kalibrierung der Kameras, das heißt der optische Fehler beider Kamerabilder und die Ausrichtung der beiden Kameras zueinander bestimmt und die entsprechenden Bilder angepasst werden. Durch Finden von Korrespondenzpunkten in beiden Bildern kann dies geschehen. Das Verfahren zur Stereokalibrierung durchläuft drei Stationen, welche in den folgenden Kapiteln näher erläutert werden.

1. **Entzerrung:** Beide Bilder können mit radialer und tangentialer Lin-
senverzerrung aufgenommen sein, welche die aufgenommene Sze-
ne verzeichnet. Diese Verzerrungen müssen zur Weiterverarbeitung
möglichst gut entfernt werden. Dieser Schritt ist vom Stereosystem
entbunden und für jede Kamera unabhängig von den anderen Kame-
ras zu berechnen.
2. **Rektifizierung:** Der Beobachtungsbereich der Kameras überschnei-
det sich größtenteils, dennoch unterscheiden sich die aufgenomme-
nen Bilder aufgrund von Position und Ausrichtung der Kameras im

Detail. Bildpunkte im rechten Bild liegen an anderen Stellen als der entsprechende Punkt im linken Bild. Um die Korrespondenzanalyse zu vereinfachen, werden die Winkel und Abstände auf Basis der Epipolargeometrie (siehe Kapitel 2.4) zwischen beiden Bildern angepasst. Korrespondierende Punkte liegen dann in beiden Bildern auf der gleichen Zeile.

3. **Korrespondenzanalyse:** Punkte im rechten und linken Bild, die den gleichen Ort im Raum präsentieren, müssen lokalisiert werden.

Ziel der Stereokalibrierung ist es, ein einheitliches Koordinatensystem für die beiden monokularen Kameras zu bilden und ihre intrinsische sowie extrinsischen Parameter derart anzupassen, dass sie die gleiche Sichtweise auf die Szene haben, um die Korrespondenzsuche zu unterstützen. Dadurch können Disparitätswerte bestimmt werden, aus denen wiederum Tiefeninformationen gewonnen werden können.

Die Kalibrierung wird unter Hinzunahme eines Schwarz/Weiß-Schachbrettes vergliche Abbildung 2.1 durchgeführt. Das wiederkehrende und einfache Muster des Schachbrettes erlaubt eine genaue Korrespondenzanalyse und somit eine präzise Kalibrierung. Die Kalibrierung der Kameras erfolgt dabei über die Suche nach inneren markanten Ecken und Kanten im Schachbrett. Durch die Kenntnis der Eck-Koordinaten innerhalb des aufgenommenen und des realen Schachbrettes kann die Orientierung und die Eigenschaften der Kamera bestimmt werden. Aus der Beziehung der Eckpunktpositionen im aufgenommenen Schachbrett und der entsprechenden Eckkoordinaten im realen Schachbrett können relevante und unabdingbare Eigenschaften und Parameter der einzelnen Kameras abgeleitet werden. Die Beschränkung der Farben des Schachbrettes auf weiß und schwarz maximiert den Kontrast an den Quadratübergängen. Die inneren Ecken sind somit sehr einfach und schnell zu finden. Ungenauigkeiten in der Koordinatenbestimmung aufgrund von Farbähnlichkeiten (zum Beispiel blau und grün) werden reduziert, sodass eine genauere Kalibrierung unterstützt wird.

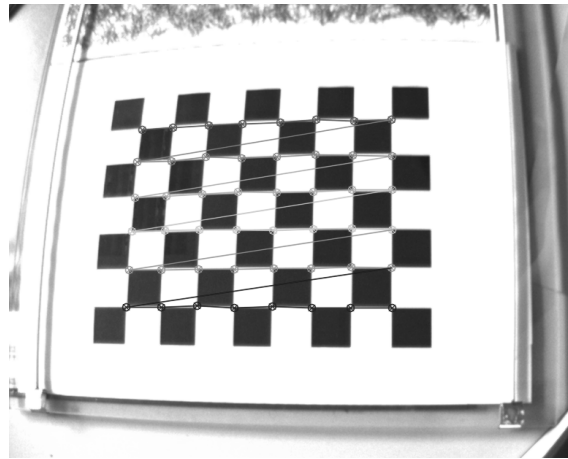


Abbildung 2.1: *Schachbrett als Kalibrierungsobjekt: Das Bild stellt ein Schachbrett mit 8x5 Quadraten dar. Es existieren somit $7 \times 4 = 28$ innere Ecken, die gefunden und lokalisiert werden müssen. In dieser Abbildung werden die gesuchten Ecken farblich markiert.*

2.1 Lochkameramodell

Das Lochkameramodell ist eine geometrische Vereinfachung komplexer optischer Systeme und beschreibt das einfachste Kameramodell. Es enthält keine optischen Linsen und charakterisiert somit Lichtstrahlen als eine Gerade vom Objekt bis auf die Bildebene. Dadurch entstehen keine Verzerrungen oder ähnliche „Verfälschungseffekte“ auf der Bildebene, die durch optische Linsen verursacht werden.

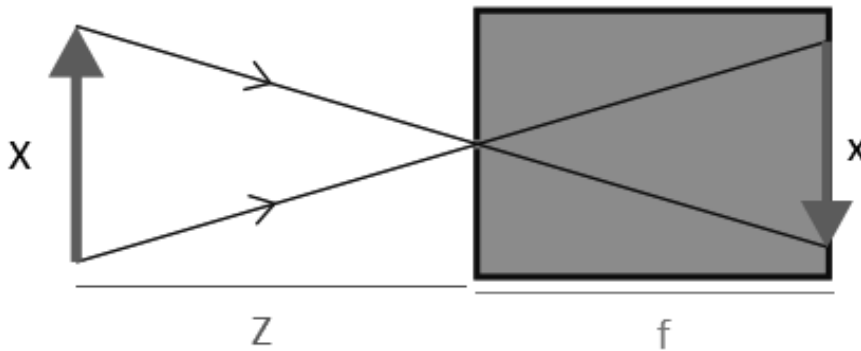


Abbildung 2.2: Die Lochkamera: Die Abbildung zeigt das Prinzip der Lochkamera. Aufgrund der fehlenden Linse fallen die Strahlen als Geraden auf die Bildebene. Es treten keine Verzerrungseffekte durch die Linse auf. Abgesehen von der seitenverkehrten Pfeilrichtung und Größe sind die beiden Pfeile gleich [18].

Wie in Abbildung 2.2 zu sehen ist, wird der links dargestellte Pfeil ohne Verzeichnungen abgebildet, sodass die Abbildung x des Pfeils X abgesehen von der negativen Pfeilrichtung und der Größe, die vom Abstand Z und von der Brennweite f abhängig ist, identisch ist. Zur Vermeidung der Seitenverkehrung der Pfeilrichtung wird die Bildebene zwischen die Projektionsfläche und dem Objekt beziehungsweise der beobachteten Szene gesetzt wie in Abbildung 2.3 dargestellt.

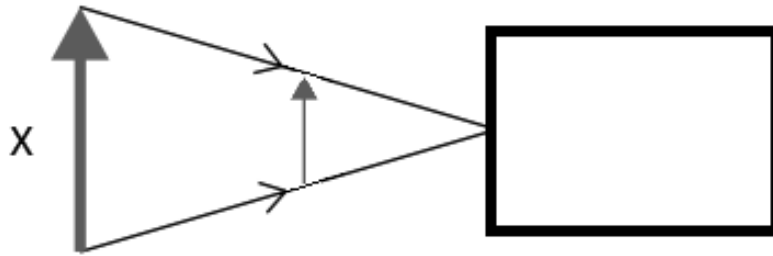


Abbildung 2.3: Das Lochkammermodell: Die Seitenverkehrung der Pfeilrichtung wurde aufgehoben. Die Bildebene wird zwischen die Projektionsfläche und dem Objekt geschoben [18].

Um die Weltkoordinaten $(X, Y, Z)^T$ in Koordinaten der Bildebene $(u, v)^T$ umzuwandeln, müssen die Ursprungskoordinaten gedreht und verschoben werden, das heißt die Transformation von Weltkoordinaten in das Kamerakoordinatensystem erfolgt über eine Rotation und eine Translation wie in Abbildung 2.4 deutlich wird.

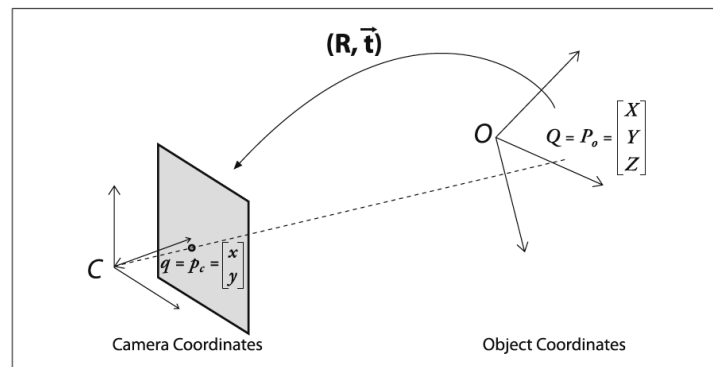


Abbildung 2.4: Perspektivische Transformation des Welt- ins Kamerakoordinatensystem: Mittels einer geeigneten Rotation und Translation lassen sich die Weltkoordinaten $P_0 = (X, Y, Z)^T$ in Kamerakoordinaten $P_C = (x, y)^T$ umwandeln [2].

Die Rotation und Translation, R und T , definieren die extrinsischen Parameter der Kamera und repräsentieren somit die äußere Orientierung der Kamera. Sie beinhalten Informationen über die Position der Kamera und die Richtung der Kamera bezüglich des gegebenen Weltkoordinatensystems. Die orthogonale Rotationsmatrix R hat drei Freiheitsgrade, die den Drehwinkeln um die drei Achsen des Koordinatensystems entsprechen. Der Translationsvektor T hat ebenfalls drei Freiheitsgrade, sodass sechs Parameter zu bestimmen sind [3].

Als Gleichung für die Transformation der 3D-Weltkoordinaten eines Punktes P_W in 3D-Kamerakoordinaten des Punktes P_C erhält man:

$$P_C = RP_W - T \quad \text{mit } P_W = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \text{ und } P_C = \begin{pmatrix} U \\ V \\ W \end{pmatrix} \quad (2.1)$$

Die Weltkoordinaten des Punktes P_W werden in die 3D-Kamerakoordinaten des Punktes P_C transformiert. P_C muss allerdings nun noch auf 2D-Sensorkoordinaten $P_B = (u, v)^T$ abgebildet werden, welche die Koordinaten auf der Bildebene definiert. Es ist eine weitere perspektivische Transformation nötig. Diese verschiebt den Hauptpunkt (c_x, c_y) , dem Schnittpunkt der optischen Achse mit der Bildebene, in den Ursprung des Bildkoordinatensystems, da der Ursprung im Normalfall in einer Ecke des Bildes liegt. Unter Hinzunahme des Verhältnisses $\frac{u}{U} = \frac{v}{V} = \frac{f}{W}$, das zwischen einem Punkt $P_C = (U, V, W)^T$ im 3D-Kamerakoordinatensystem und dem transformierten Punkt $P_B = (u, v)^T$ im 2D-Kamerakoordinatensystem besteht, lässt sich die Projektionsmatrix darstellen durch die sogenannte Kameramatrix

$$M = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad \text{mit } \begin{matrix} f_x, f_y & \text{Brennweiten} \\ (c_x, c_y) & \text{Hauptpunkt} \end{matrix} \quad (2.2)$$

Diese beinhaltet die intrinsischen Parameter der Kamera und charakterisiert somit die interne Geometrie der Kamera. Sie enthält die Brennweiten f_x und f_y in x- und y-Richtung, sowie die Pixelkoordinaten des Hauptpunktes (c_x, c_y) . Mathematisch wird die Umwandlung von 3D-Koordinaten im Weltkoordinatensystem $P_W = (X, Y, Z)^T$ in 2D-Bildkoordinaten $P_B = (u, v)^T$ durch eine perspektivische Transformation zusammenfassend folgendermaßen modelliert:

$$P_B = [MRT]P_W \quad (2.3)$$

$$\Leftrightarrow \begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad (2.4)$$

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f_x & 0 \\ 0 & f_y \end{pmatrix} \begin{pmatrix} \frac{U}{W} \\ \frac{V}{W} \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}$$

Zur Bestimmung der extrinsischen und der intrinsischen Parameter, R, T und M, werden mehrere Schachbrett-Aufnahmen der Kamera benötigt. Das Schachbrett wird aus verschiedenen Perspektiven aufgenommen. Dabei ist darauf zu achten, dass möglichst das gesamte Schachbrett in jeder Perspektive zu sehen ist. Ansonsten können nicht alle inneren Ecken exakt erkannt und lokalisiert werden. Anschließend werden die inneren Eckpunkte der Quadrate in den aufgenommenen Schachbrett-Bildern gesucht. In Abbildung 2.1 sind die gesuchten Eckpunkte innerhalb des Schachbrettes eingezeichnet. Die gefundenen Koordinaten der Eckpunkte im aufgenommenen Schachbrett werden im Folgenden als Bildpunkte bezeichnet. Die Positionen der entsprechenden Koordinaten im realen Schachbrett sind ausmessbar. Diese werden Objektpunkte genannt. Anhand der Bild- und Objektpunkte kann die obige Formel durch ein lineares Gleichungssystem dargestellt werden. Durch Lösen dieses Gleichungssystems können die gesuchten Kameraspezifikationen approximiert werden.

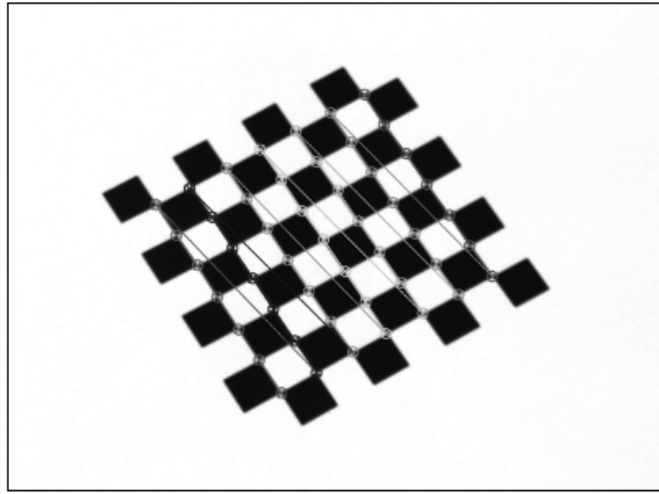


Abbildung 2.5: *Bestimmung der internen Eckpunkte: Im Beispielbild existieren $7 \times 7 = 49$ interne Ecken, welche gesucht werden müssen. Diese wurden farblich markiert. Wenn sie lokalisiert werden können, ist eine Approximation der Kameramatrix M und der extrinsischen Parameter R und T möglich.*

Eine Aufnahme des Schachbrettes reicht nicht aus, um eine stabile und robuste Lösung des Gleichungssystems zu erreichen. Durch mehrere Aufnahmen aus unterschiedlichen Perspektiven ist es möglich, eine genauere Approximation von Kameramatrix, Rotationsmatrix und Translationsvektor zu gewährleisten, da mehr Korrespondenzpunkte und somit mehr Gleichungen zur Parameterberechnung vorliegen.

Durch die gegebenen Korrespondenzen zwischen Bild- und Objektpunkten lässt sich ein nichtlineares Gleichungssystem aufstellen, was durch ein nichtlineares Optimierungsverfahren die extrinsischen und intrinsischen Parameter der Kamera abschätzt. Insgesamt sind zehn Unbekannte zu berechnen: Sechs Parameter beschreiben die drei Rotationswinkel und die Verschiebungen in die drei Dimensionen. Vier Parameter sind für die intrinsischen Parameter (Brennweiten (f_x, f_y) und Hauptpunkt (c_x, c_y)) zu

bestimmen. Es sind somit mindestens zehn Korrespondenzpunkte nötig, um das Gleichungssystem zu lösen. Dennoch gilt: Je mehr Korrespondenzpunkte zur Verfügung stehen, desto mehr Gleichungen können aufgestellt werden, um eine genauere Approximation der gesuchten Parameter zu garantieren. Störungen und Inkonsistenzen können reduziert und gegebenenfalls eliminiert werden. Ein Algorithmus zur Bestimmung des Gleichungssystems stammt von Zhang [7], welcher die Eigenschaften des planaren Kalibrierungsobjektes und die Orthogonalität der Rotationsmatrix R ausnutzt [12]. Dabei wird zunächst eine Startlösung gesucht, um anschließend mittels eines nichtlinearen Optimierungsverfahrens gemäß der Minimierung des quadratischen Fehlers die gesuchten Parameter zu optimieren.

Die Kamera- und Rotationsmatrix sowie der Translationsvektor werden zu einer 3×3 Homographiematrix H zusammengefasst, die damit alle wesentlichen Parameter beinhaltet. Diese beschreibt dann letztendlich die Projektion der Welt- in Bildkoordinaten [3]. Da das Kalibrierungsobjekt planar ist, ist keine Rotation in z -Richtung nötig. Diese wird daher in der Homographiematrix vernachlässigt.

$$\begin{aligned}
 H &= MRT \\
 &= \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{pmatrix} \\
 &= \begin{pmatrix} f_x r_{11} + c_x r_{31} & f_x r_{12} + c_x r_{32} & f_x t_1 + c_x t_3 \\ f_y r_{21} + c_y r_{31} & f_y r_{22} + c_y r_{32} & f_y t_2 + c_y t_3 \\ r_{31} & r_{32} & t_3 \end{pmatrix}
 \end{aligned} \tag{2.5}$$

Ein eindeutiger Zusammenhang zwischen 3D-Raumkoordinaten und seiner Abbildung auf die 2D-Bildebene in Pixelkoordinaten lässt sich damit herstellen [3]. Diese Beziehung wird bei der Entzerrung eine tragende Rolle spielen, da sie zur Approximation der relevanten Kalibrierungsparameter und der Verzerrungskoeffizienten notwendig ist.

2.2 Verzerrungen

Im Gegensatz zum Modell der Lochkamera existieren bei üblichen Kameras durch die Linsen verursachte negative Effekte. Obwohl die heutige Technik immer weiter fortschreitet, gibt es keine Kameralinse, die perfekt ist. Kameralinsen sind mit Störungen und Verzerrungen behaftet. Selbst mit einem sehr hohen Kostenaufwand sind diese Störungen und Verfälschungen nahezu unvermeidbar.

Bei den Verzerrungsmodellen existieren zwei, die relevant für die Kamerakalibrierung sind:

1. Radiale Verzerrung
2. Tangentiale Verzerrung

2.2.1 Radiale Verzerrung

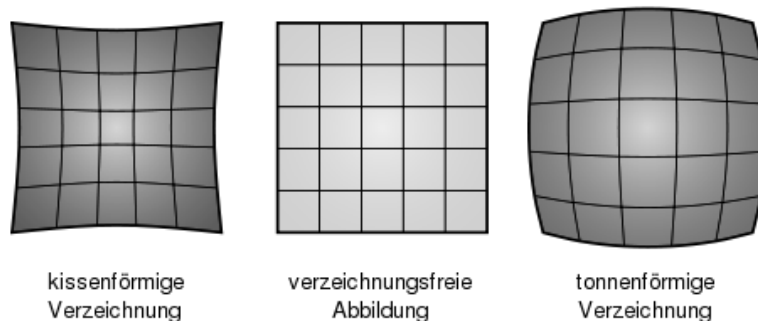


Abbildung 2.6: *Beispiel für radiale Verzerrung: Das unverzerrte Gitternetz in der Mitte wird durch die radiale Verzerrung abgeändert. Es kann eine kissen- (links) oder tonnenähnliche (rechts) Form annehmen [12].*

Die radiale Verzerrung wird durch die Kameralinse verursacht. Dieser Effekt spiegelt sich in den Bildern in der Weise wieder, dass die Kanten im

Bild gekrümmt dargestellt und die Ecken abgerundet oder spitzer werden vergleiche Abbildung 2.6. Geraden werden nicht mehr auf Geraden abgebildet.

In der Mitte des Bildes ist die radiale Verzerrung nicht vorhanden, da an dieser Stelle die Kameralinse noch eben ist. Doch je mehr sich die Linse wölbt, desto mehr wird der Effekt der radialen Verzerrung deutlich. Abbildung 2.7 zeigt, wie die radiale Verzerrung mit dem Radius wächst. In

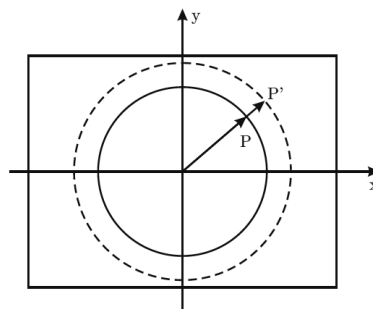


Abbildung 2.7: Prinzip der radialen Verzerrung: Der Punkt P wird entlang des Radius auf den Punkt P' abgebildet. Je größer der Radius wird, desto weiter oder näher wird der Punkt verschoben [16].

der Optik wird die radiale Verzerrung durch eine Funktion abhängig vom Abstand zum Mittelpunkt der Linse, welcher die Punkte längs deren radialen Richtungen verschiebt, modelliert. Die additive Änderung des Radius durch diese Verschiebung ist durch ein Polynom zweiter Ordnung approximierbar:

$$v(r) = 1 + k_1 r^2 + k_2 r^4 + k_3 r^6 \text{ mit } r^2 = x^2 + y^2 \quad (2.6)$$

Als Annäherung reichen bereits die Koeffizienten k_1 , k_2 und k_3 , da mit steigender Potenz vom Radius r die Werte immer weniger Einfluss auf den Verzerrungseffekt haben, das heißt die radiale Verzerrung ist durch ein Polynom der Ordnung 2 abschätzbar.

Die korrekten Bildkoordinaten (x, y) können anschließend mittels der radialen Verzerrungsfunktion und den radial verzerrten Bildkoordinaten (x_r, y_r) berechnet werden [12].

2.2.2 Tangentiale Verzerrung

Die tangentiale Verzerrung entsteht durch die Bauart der Kamera, wenn der Hauptsehstrahl nicht senkrecht durch die Linse fällt und die Linse leicht zu ihm geneigt ist. Deutlich wird die tangentiale Verzerrung durch den gleichzeitigen Auftritt von tonnen- und kissenförmigen Effekten in den Kameraaufnahmen vergleiche Abbildung 2.8. Die Abbildung 2.9 zeigt,

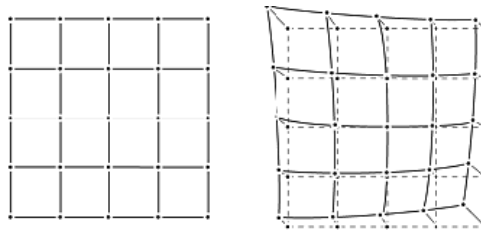


Abbildung 2.8: *Beispiel für die tangentiale Verzerrung: Das unverzerrte rechteckige Gitternetz (links) wird durch die tangentiale Verzerrung (rechts) verzeichnet. Der kissen- und tonnenförmige Effekt aus der radialen Verzerrung tritt gleichzeitig auf. Daher ist die tangentiale Verzerrung auch unter dem Begriff der radial asymmetrischen Verzeichnung bekannt [12].*

dass die tangentiale Verzeichnung einen Punkt entlang der Tangenten an den angrenzenden Kreis verschiebt, welcher um das Linsenzentrum durch den abzubildenden Punkt definiert wird. Diese Verschiebung ist keine Funktion des Radius, sondern eine Verknüpfung der Verzerrungskoeffizienten p_1 und p_2 mit dem Bildpunkt $P = (x, y)$. Wie die radiale, wächst die tangentiale Verzerrung ebenfalls mit dem Abstand vom optischen Zentrum. Die tangentiale Verzerrung kann nicht mehr nur als Funktion des Bildradius modelliert werden, sondern ist ebenfalls von den tangentialen Verzer-

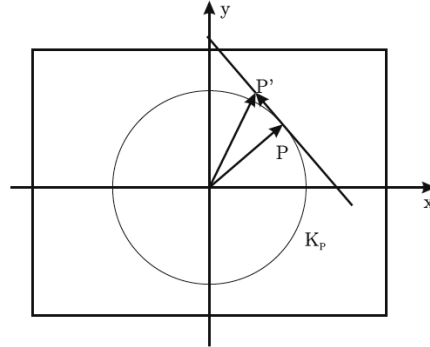


Abbildung 2.9: *Prinzip der tangentialen Verzerrung: Der Punkt P wird entlang der Tangenten an den Kreis K_P auf den Punkt P' versetzt [16].*

rungsfaktoren p_1 und p_2 , sowie von den Koordinaten eines betrachteten tangential verzerrten Bildpunktes $P = (x_t, y_t)$ abhängig [3]:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} 2p_1x_ty_t + p_2(r^2 + 2x_t^2) \\ 2p_2x_ty_t + p_1(r^2 + 2y_t^2) \end{pmatrix} \quad (2.7)$$

Dabei ist $P' = (x_t, y_t)$ der verzerrte abgebildete Punkt.

2.3 Entzerrung

Um die Verzerrung durch die Kameralinse zu korrigieren, müssen die aufgenommenen Bilder durch Methoden der Bildverarbeitung entzerrt werden. Dabei nutzt man das Modell der Lochkamera und modelliert die Entzerrung unter Erweiterung der Gleichung 2.3.

Um nun die Welt- in Kamerakoordinaten trotz Verzeichnungen umzuwandeln, reichen die intrinsischen und extrinsischen Parameter für die Transformation nicht mehr aus. Auch die Verzerrungskoeffizienten k_1, k_2, k_3, p_1

und p_2 spielen nun eine relevante Rolle. Das Lochkameramodell wird modifiziert dargestellt durch:

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad (2.8)$$

$$U' = \frac{U}{W} (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 \frac{U}{W} \frac{V}{W} + p_2 (r^2 + 2 \frac{U^2}{W^2}) \quad (2.9)$$

$$V' = \frac{V}{W} (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_2 \frac{U}{W} \frac{V}{W} + p_1 (r^2 + 2 \frac{V^2}{W^2}) \quad (2.10)$$

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f_x & 0 \\ 0 & f_y \end{pmatrix} * \begin{pmatrix} U' \\ V' \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix} \quad (2.11)$$

Die Koordinaten (U', V') sind die verzerrten Koordinaten nach der Transformation der 3D-Weltkoordinaten in das 3D-Kamerakoordinatensystem, die durch Anwenden der Verzerrungsfunktionen 2.6 und 2.7 verzeichnet werden. Diese werden anschließend durch Multiplikation mit der Kameramatrix in Bildkoordinaten umgewandelt.

Um die relevanten Kameraparameter einer herkömmlichen Kamera zu bestimmen, geht man zunächst von einer idealen Kamera aus. Man nutzt das Lochkameramodell und berechnet mit dem in Kapitel 2.1 vorgestellten Algorithmus von Zhang [7] eine erste Schätzung für die gesuchten Parameter. Mehrere Aufnahmen des Schwarz/Weiß-Schachbrettes aus verschiedenen Perspektiven dienen zur Lokalisierung von Bild- und Objektpunkten, die durch die inneren Eckpunkte bestimmt werden (siehe Kapitel 2.1). Bei der Implementierung der Entzerrung wurde mit Kameraaufnahmen aus fünf bis zehn unterschiedlichen Blickwinkeln gearbeitet. In jeder Aufnahme war das ganze Schachbrett zu sehen. Ein Beispiel für ein Testbild ist Abbildung 2.10. Dabei werden bei der projektivischen Transformation der

Welt- ins Kamerakoordinatensystem beim Lochkameramodell die Verzerrungskoeffizienten auf 0 gesetzt. Anschließend werden über die Gleichungen 2.9 und 2.10 die unbekannten Verzerrungskoeffizienten approximiert. Durch diese kann eine Überprüfung der vorher geschätzten extrinsischen und intrinsischen Parameter erfolgen. Diese werden bei einem zu hohen quadratischen Fehler neugeschätzt bis sich der Fehler nicht mehr reduzieren lässt und somit minimiert wurde [3].

Da es sich bei den vorgestellten Verzerrungseffekten nur um Modelle handelt, sind die Verzeichnungen nie komplett entfernbar. Sie können nur minimiert werden, sodass die vorhandenen Verzeichnungen die Weiterverarbeitung und Analyse der Bilder möglichst wenig beeinträchtigen.

Die beiden folgenden Abbildungen 2.10 und 2.11 zeigen die Entzerrung an einem Beispiel auf. Dargestellt wird eine Kameraaufnahme eines Schachbrettes, die erhebliche Verzerrungen ausweist. Diese spiegeln sich in den gewölbten Kanten und abgerundeten Ecken des Schachbrettes wider. Die Verzeichnungen müssen minimiert und soweit wie möglich reduziert werden, um die weitere Verarbeitung und Auswertung der Kameraaufnahme zu vereinfachen. Durch die gefundenen Ecken des Schachbrettes sind so-

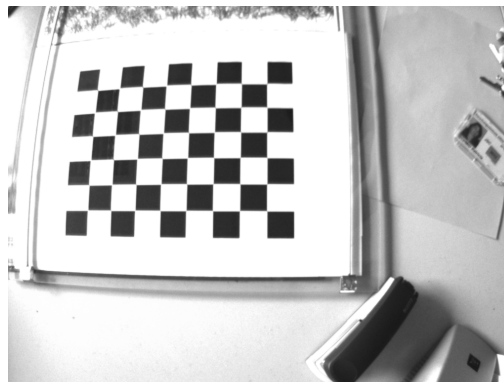


Abbildung 2.10: *Verzerrtes Bild: Das Beispielbild zeigt ein Schachbrett, welches mit einer monokularen Kamera aufgenommen wurde. Es enthält erhebliche Verzerrungen, die sich in den gewölbten Kanten und abgerundeten Ecken bemerkbar machen.*

wohl Bild- als auch Objektpunkte bekannt. Die benötigten Verzerrungskoeffizienten können approximiert werden. Dadurch ist eine Minimierung der Verzerrungen möglich, sodass die Analyse der Aufnahme unproblematisch wird.

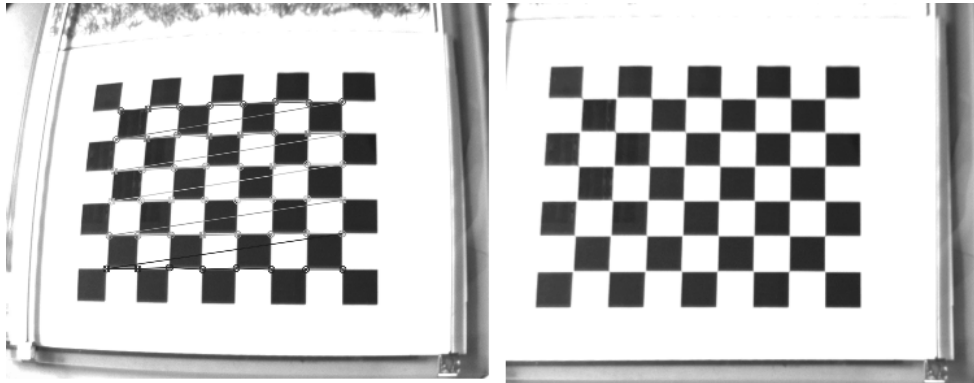


Abbildung 2.11: *Entzerrtes Bild: Alle inneren Ecken wurden im verzerrten Bild gefunden. Diese sind farblich markiert dargestellt. Das Bild konnte dadurch und mittels des vorgestellten Algorithmuses erfolgreich entzerrt werden.*

2.4 Epipolargeometrie

Die Epipolargeometrie ist ein mathematisches Modell zur Darstellung geometrischer Beziehungen zwischen verschiedenen Aufnahmen des gleichen Objektes. Die Kamerabilder werden aus unterschiedlichen Perspektiven aufgenommen. Die Epipolargeometrie beschreibt dabei die Abhängigkeiten und geometrischen Zusammenhänge zwischen den korrespondierenden Bildpunkten [20]. Sie kombiniert zwei Lochkameramodelle, sodass eine vorherige Entfernung beziehungsweise Minimierung der vorhandenen Verzerrung unabdingbar ist. Erst ohne Verzeichnungen lassen sich die verwendeten Kameramodelle auf das der Lochkamera übertragen.

Wie schon in Kapitel 2.1 erwähnt, lässt sich eine Kamera nach der Entzerrung geometrisch durch eine Lochkamera modellieren. Bei dieser liegt während der Aufnahme jeder Punkt des aufgenommenen Objektes, das Projektionszentrum (entspricht dem Linsenzentrum der Kamera) sowie der dazugehörige Bildpunkt auf einer Geraden. Wurde der Bildpunkt zweimal von unterschiedlichen Positionen aus aufgenommen, lassen sich bei einer späteren Auswertung mittels der Orientierungen der Kameras der Schnittpunkt der zwei Geraden und damit die Koordinaten des beobachteten Punktes berechnen. Eine 3D-Rekonstruktion ist somit möglich, wenn in beiden Bildern die korrespondierenden Bildpunkte lokalisiert wurden. Die Epipolargeometrie dient zur Unterstützung dieser Lokalisation: Ist der Punkt im ersten Bild gegeben, schränkt sich bei bekannter Epipolargeometrie der Suchbereich im zweiten Bild auf eine Linie, der Epipolarlinie, ein [13].

Für die Epipolargeometrie existiert eine feste Terminologie, welche in Abbildung 2.12 dargestellt wird [20].

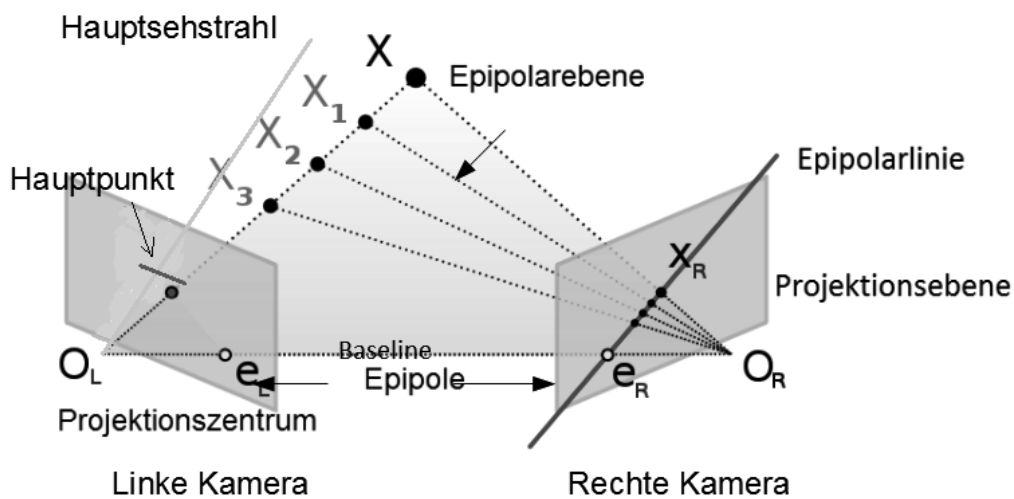


Abbildung 2.12: *Epipolargeometrie: Es wird die geometrische Beziehung zwischen zwei Kameras mittels der Epipolargeometrie dargestellt. Die feste Terminologie der Epipolargeometrie wird verdeutlicht und in unten stehender Aufzählung erläutert [20].*

1. **Projektionsebene** definiert die Bildebene.
2. **Projektionszentrum** ist das Zentrum der Kamera (hier: O_L und O_R).
3. **Baseline** beschreibt die Gerade durch die Zentren der beiden Kameras O_R und O_L .
4. **Hauptsehstrahl** ist der Projektionsstrahl, welcher durch das Projektionszentrum und senkrecht auf der Projektionsebene steht.
5. **Hauptpunkt** definiert den Schnittpunkt der Bildebene mit dem Hauptsehstrahl. Er ist die orthogonale Projektion des Projektionszentrums auf die Bildebene.
6. **Epipol** ist der Schnittpunkt der Baseline mit der Projektionsebene (hier: e_L und e_R).
7. **Epipolarebenen** sind alle Ebenen, welche die Baseline enthalten.

8. **Epipolarlinie** beschreibt die Linie, die beim Schnitt der Projektionsebene mit der Epipolarebene entsteht.

Verdeutlicht wird die Epipolargeometrie an einem Beispiel: Uns stehen zwei Kameras zur Verfügung, die wie in Abbildung 3.4 ausgerichtet sind. Ihre Projektionszentren liegen bei O_L und O_R . Alle anderen geometrischen Beziehungen verhalten sich ebenfalls wie in obiger Abbildung. Beobachtet man nun einen Punkt X_L in der linken Bildebene, kennt man die Tiefe des Punktes in der Regel nicht. Es kommen alle Punkte in Frage, die auf der durch X_L und O_L liegenden Geraden der Epipolarebene liegen. Diese sind zum Beispiel die Punkte X_1 , X_2 und X . Bekannt ist lediglich, dass der korrespondierende Punkt auf einer Schnittgeraden der Epipolarebene mit der Projektionsebene, der Epipolarlinie, der rechten Kamera liegen muss. Der Suchbereich eines korrespondierenden Punktes lässt sich folglich auf diese Linie einschränken. Findet man auf dieser Epipolarlinie den korrespondierenden Punkt, hier X_R , kann aufgrund der Geometrie auch die Tiefe des beobachteten Punktes ermittelt werden. Der “3D-Punkt“ ist dann der Schnittpunkt X zwischen den beiden Projektionsstrahlen durch die Punkte X_L und X_R .

Die Epipolargeometrie kombiniert zwei Lochkameras, sodass vorhandene Verzerrungen vorher entfernt werden müssen. Die beiden Kameras sind verschoben und zueinander gedreht. Durch eine geeignete Koordinatentransformation kann das Weltkoordinatensystem in das optische Zentrum der ersten (hier: linken) Kamera gelegt werden, das heißt das Koordinatensystem der zweiten (hier: rechten) Kamera wird, wie in Kapitel 2.1 beschrieben und in Abbildung 2.13 zu sehen, in das linke Kamerakoordinatensystem transformiert, was wieder durch eine Rotation und Translation geschieht [2].

$$P_l = RP_r - T \quad (2.12)$$

P_r symbolisiert die Punkte im rechten Bild, welche durch Transformation in das erste Kamerakoordinatensystem gelegt werden. P_l stellt somit

die Punkte P_r bezüglich des ersten Koordinatensystems dar und beschreibt diese in Koordinaten des linken Bildes.

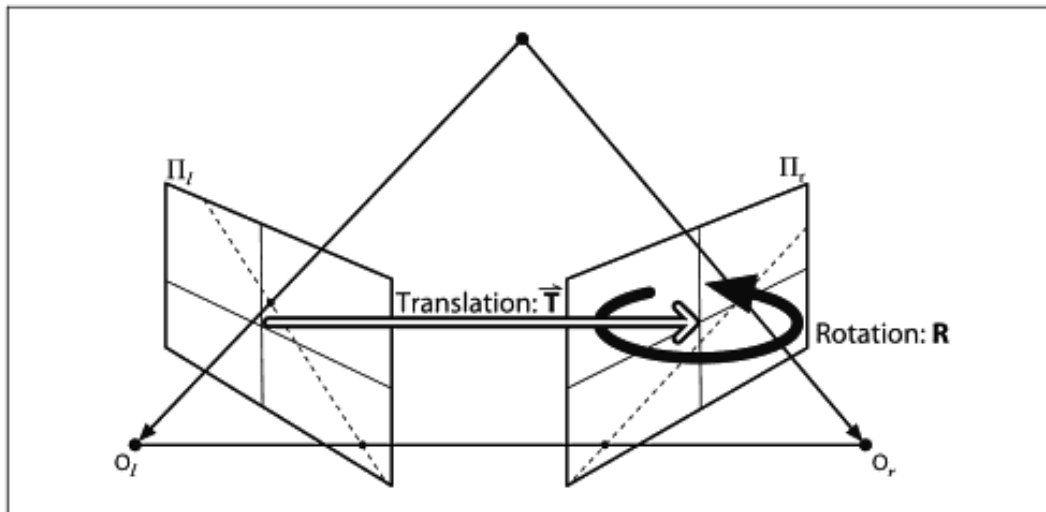


Abbildung 2.13: *Koordinatentransformation der Epipolargeometrie: Durch eine Drehung und Verschiebung ist eine Transformation des linken Koordinatensystem in das rechte Koordinatensystem möglich. Mit anschließender Verknüpfung der Transformationsvorschrift mit den Kameramatrizen beider Kameras erfolgt die perspektivische Transformation. Eine mathematische Beziehung der Positionen beider Kameras und damit auch ihrer Bildpunkte zueinander ist geschaffen [2].*

Die perspektivische Transformation zwischen den Bildkoordinaten korrespondierender Punkte wird durch die Fundamentalmatrix F definiert. Mit ihr lässt sich zu einem gegebenen Punkt im ersten Bild die dazugehörige Epipolarlinie im zweiten Bild bestimmen, auf der sich der korrespondierende Punkt befindet [3]. Die Fundamentalmatrix und damit die Epipolargeometrie lässt sich bei bekannter Kalibrierung der beiden Kameras direkt aus den intrinsischen und extrinsischen Parametern beider Kameras sowie einem Projektionszentrum berechnen [20].

$$F = (M_r)^{-T} T_x R (M_l)^{-1} \text{ mit } T_x R = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad (2.13)$$

Mit Hilfe der Epipolargeometrie kann die Korrespondenzanalyse vereinfacht werden. Ohne Verwendung der Epipolargeometrie und ihrer Eigenschaften müssen jeweils die ganzen Bilder nach dem Korrespondenzpunkt durchsucht werden. Ist die Epipolargeometrie bekannt, können die Kameraaufnahmen rektifiziert werden (siehe Kapitel 2.5), damit korrespondierende Punkte nicht nur auf gleicher Epipolarlinie, sondern auch auf gleicher Zeile liegen. So ist der Suchbereich auf eine Epipolarlinie beschränkt. Durch den erheblich verringerten Suchaufwand kann an Rechenzeit gespart und die Zuverlässigkeit erhöht werden.

2.5 Rektifizierung

Das Problem bei einem Stereosystem ist, dass die Kameras nie in einer Ebene liegen. Selbst bei dem Versuch sie so genau wie möglich in gleicher Höhe zu platzieren, enthalten die Bildaufnahmen deutliche Abweichungen nicht nur in horizontaler Richtung. Um Korrespondenzen von gleichen Punkten zu finden, ist es als Vorverarbeitungsschritt sinnvoll die Bilder auszurichten. Deswegen braucht man das Verfahren der Rektifizierung. Bei der Rektifizierung werden die Bilder mittels der Epipolargeometrie so reprojeziert, dass sie parallel konfiguriert sind und in der gleichen Ebene liegen, das heißt, die Kameras werden so ausgerichtet, dass sich Zeilen entsprechen und sich korrespondierende Punkte nicht nur auf gleicher Epipolarlinie, sondern auch auf gleicher Zeile befinden.

Wenn Punkte in einem Bild auf der gleichen Zeile liegen wie die entspre-

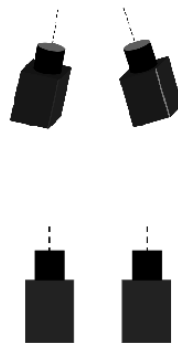


Abbildung 2.14: *Rektifiziertes Stereokamerasystem: Beide monokularen Kameras haben eine bestimmte Orientierung zueinander (oben), das heißt durch eine geeignete Rotation und Translation ist die eine Kamera in die Position der anderen Kamera transformierbar. Durch die Rektifizierung mittels der bekannten Epipolargeometrie kann die perspektivische Transformation berechnet und die beiden Kameras zueinander so ausgerichtet werden, dass sie parallel zueinander liegen [16].*

chenden Punkte im anderen Bild, wird der nächste Schritt, die Korrespondenzsuche, vereinfacht. Dadurch muss bei der Korrespondenzanalyse nicht

mehr das gesamte Bild durchlaufen werden, sondern nur noch die entsprechende Bildzeile muss durchsucht werden.

Zur Rektifizierung muss die Epipolargeometrie der beiden Kameras zueinander bekannt sein. Die geometrischen Beziehungen und Eigenschaften sind unabdingbar. Anhand der abgeschätzten Fundamentalmatrix können die Winkel und Abstände im Bild angepasst werden und die Epipolarlinien damit parallelisiert werden.

Wenn die Epipolarlinien parallel zueinander liegen, sind die Kameras wie in Abbildung 2.14 ebenfalls parallel zueinander ausgerichtet, beziehungsweise die aufgenommenen rektifizierten Bilder erscheinen als wären sie von exakt parallel zueinander liegenden Kameras aufgenommen. Sie sind entlang einer Bildachse verschoben sind und die Hauptsehstrahlen werden parallelisiert.

Zur Realisierung der Rektifizierung existiert ein Algorithmus von Bouguet, der die Kenntnis der intrinsischen und extrinsischen Kameraparameter ausnutzt [9]. Wegen der Epipolargeometrie und der Entzerrung sind alle benötigten Eigenschaften beider Kameras und ihre geometrischen Positionen zueinander bekannt. Die Rotationsmatrizen jeder Kamera werden berechnet, damit sie die gleiche Bildebene haben.

Die beiden folgenden Abbildungen erläutern die Rektifizierung an einem Beispiel. Die beiden Kameraaufnahmen wurden von monokularen Kameras aufgenommen, die ein Stereosystem bilden und deren Beobachtungsbereiche sich weitestgehend überlappen. Sie sind nahezu auf gleicher Höhe angebracht und unterscheiden sich nicht nur in ihrer x-Position. In den beiden Bildern sind die Verzerrungen entfernt worden. Dennoch gibt es wesentliche Unterschiede zwischen den Abbildungen im linken und rechten Bild. Sucht man nun einen Punkt aus dem linken Bild im rechten Bild, muss das gesamte rechte Bild Pixel für Pixel durchlaufen werden. Durch die Rektifizierung liegen alle korrespondierenden Punkte auf derselben Zeile. Die Korrespondenzsuche wird erheblich vereinfacht und ist nicht mehr über das gesamte rechte Bild durchzuführen. Es muss nur noch in x-Richtung nach dem entsprechenden Bildpunkt gesucht werden. Der Suchaufwand wird reduziert und somit auch an Rechenzeit gespart.

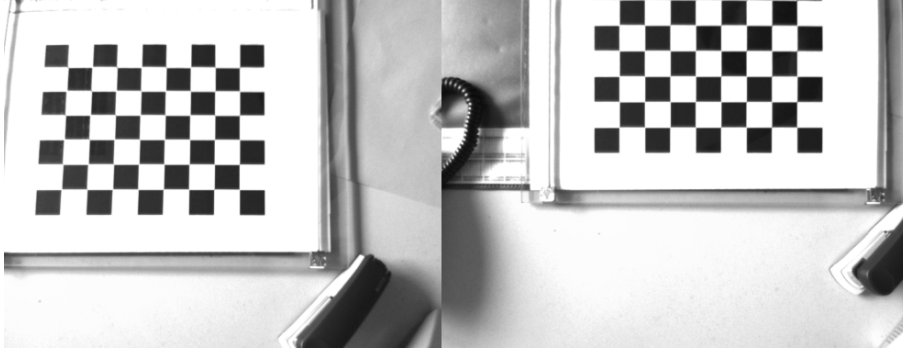


Abbildung 2.15: *Vor der Rektifizierung: Die beiden Bilder wurden von monokularen Kameras aufgenommen. In den beiden Bildern sind die Verzerrungen entfernt worden. Dennoch gibt es wesentliche Unterschiede zwischen den Abbildungen im linken und rechten Bild.*

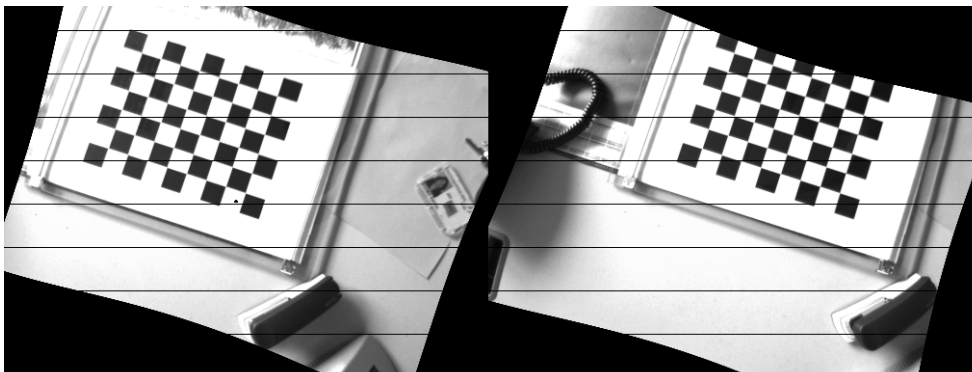


Abbildung 2.16: *Nach der Rektifizierung: Nach der Rektifizierung ist die Korrespondenzsuche über das gesamte rechte Bild nicht mehr notwendig. Alle korrespondierenden Punkte liegen auf derselben Zeile. Die eingezeichneten Linien dienen zur Visualisierung der Epipolarlinien. Sie verdeutlichen den eingeschränkten Suchbereich auf eine Zeile.*

Kapitel 3

Disparitäten

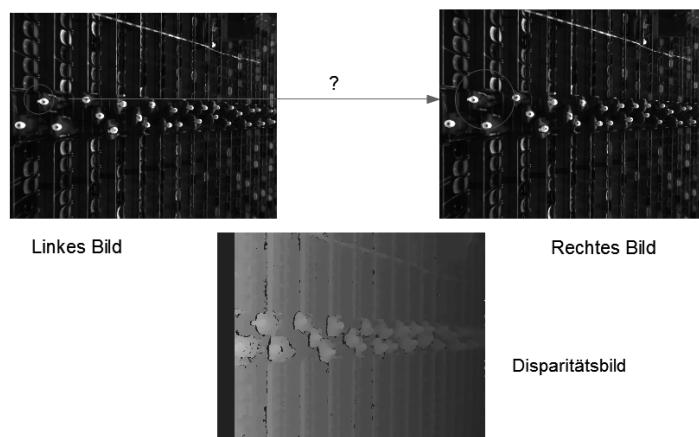


Abbildung 3.1: *Korrespondenzanalyse: Es wird ein beobachteter Bildpunkt aus dem linken Bild im rechten Bild gesucht. Die Frage ist, an welcher Stelle im rechten Bild der entsprechende Punkt liegt. Aufgrund der unterschiedlichen Perspektiven kann dieser verschoben sein und muss dementsprechend nach der Rektifizierung auf einer ganzen Zeile gesucht werden. Die Differenzen der x -Koordinaten der korrespondierenden Punkte werden in einem Disparitätsbild visualisiert.*

KAPITEL 3. DISPARITÄTEN

Die Bestimmung der Entfernung eines Objektes zur Kamera erfolgt über die Berechnung von Disparitäten. Disparität beschreibt dabei die Differenz zwischen einem Pixel im linken Bild und dem entsprechenden Bildpunkt im rechten Bild (siehe Abbildung 3.1). Aufgrund der Rektifizierung liegen korrespondierende Punkte auf der gleichen Zeile, sodass nur die Differenz in x-Richtung relevant und zu berechnen ist. Wie in Abbildung 3.2 verdeutlicht, weist eine hohe Disparität auf eine kurze Entfernung des detektierten Objektes zur Kamera hin. Dementsprechend stellt eine geringe Disparität einen weiten Abstand von Kamera zu Gegenstand dar.

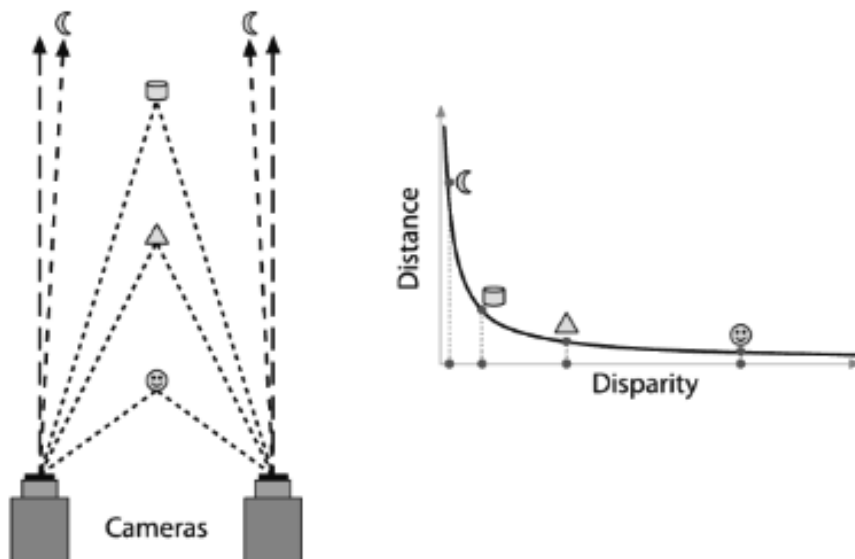


Abbildung 3.2: Zusammenhang Disparität und Abstand zur Kamera: Je weiter Objekte oder Punkte von der Kamera entfernt sind, desto kleiner ist die Differenz zwischen den x-Koordinaten des linken und rechten Bildes [2].

Mathematisch wird die Berechnung der einzelnen Disparitäten in rektifizierten Bildern durch

$$d = x_l - x_r \quad (3.1)$$

beschrieben, wobei x_l die x-Koordinate eines Punktes im linken Bild und x_r die x-Koordinate des korrespondierenden Bildpunktes im rechten Bild darstellt.

In der Abbildung 3.3 wird die Herleitung des mathematischen Zusammenhangs zwischen Disparität und Abstand eines Pixels zur Kamera aufgezeigt. Voraussetzung ist, dass die Brennweiten beider Kameras identisch sind und es gilt

$$f = f_l = f_r \quad (3.2)$$

Wendet man den Strahlensatz auf die Epipolargeometrie an

$$\frac{T - (x_l - x_r)}{Z - f} = \frac{T}{Z} \quad (3.3)$$

und formt dieses entsprechend nach dem Abstand Z um

$$Z = \frac{fT}{d} \quad \text{mit } d = x_l - x_r \quad (3.4)$$

so erhält man die mathematische Beziehung zwischen der Disparität d und dem Abstand eines Pixel zur Kamera Z . T und f definieren dabei die Länge der Baseline und die Brennweite, die in Abbildung 3.3 eingezeichnet sind. Der Abstand eines Bildpunktes zur Kamera ist somit invers proportional zur Disparität. Deutlich wird dabei, dass der Abstand eines Pixels

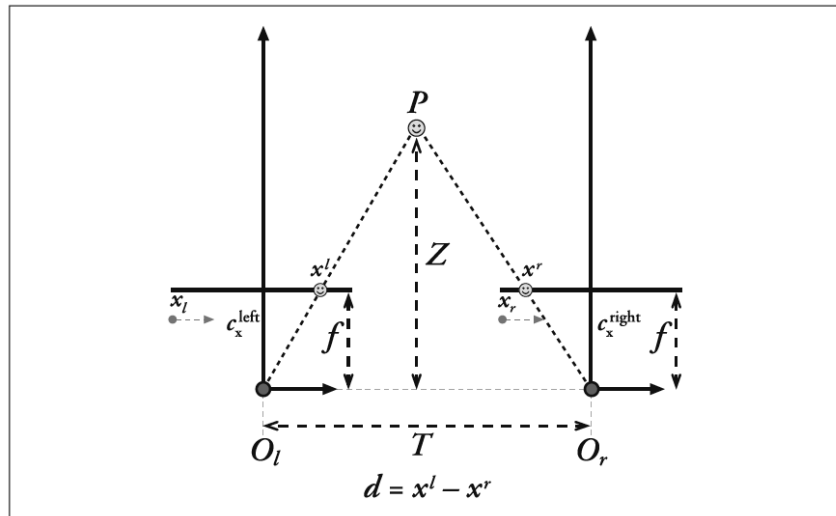


Abbildung 3.3: *Disparität und Abstand: Mittels der geometrischen Zusammenhänge der Kameraparameter und der Kameras zueinander kann der Abstand jedes Pixels zur Kamera über die Disparität ermittelt werden. Durch Anwenden des Strahlensatzes wird eine mathematische Beziehung hergestellt [2].*

zur Kamera unter Anderem abhängig von der Brennweite der Linsen und der Baseline zwischen den Kameras ist. Es zeigt sich die Motivation eine individuelle Stereokamera zu konstruieren um zum Beispiel die Baseline variabel zu halten.

Berechnet man für jeden Bildpunkt im linken Bild die Disparität zum korrespondierenden Punkt im rechten Bild, so erhält man eine Disparitätsmatrix. Zur Visualisierung der Disparitätsmatrix wird diese in einem Disparitätsbild dargestellt. Dabei werden den Differenzwerten feste Farbwerte zugeordnet, um eine repräsentative Darstellung zu gewährleisten. Die in dieser Arbeit verwandte Farbpalette beschränkt sich hierbei auf eine Grauskala, welche von schwarz nach weiß reicht. Je heller ein Pixel dargestellt wird, desto höher ist der dazugehörige Disparitätswert, das heißt bei weißen Pixeln herrscht die höchste Differenz. Dementsprechend je dunkler der

Bildpunkt ist, desto geringer ist die Disparität an dieser Stelle. Schwarz bedeutet dennoch nicht der Farbwert für die kleinste Disparität, sondern definiert den Farbwert für eine unbestimmte Disparität. An schwarzen Bildpunkten im Disparitätsbild gibt es keinen Eintrag in der Disparitätsmatrix. An diesen Stellen konnte kein Korrespondenzpunktpaar gefunden und somit keine Disparität bestimmt werden. Die Zuordnung von korrespondierenden Punkten im linken und rechten Bild ist nicht eindeutig oder kann gar nicht ermittelt werden und kann somit nicht definiert werden. Problematisch ist dieser Effekt vorallem bei texturlosen oder texturarmen Aufnahmen. Dort kann aufgrund der Gleichheit von vielen Bildpunkten und der fehlenden Struktur nicht entschieden werden, welcher Bildpunkt im linken zu welchem Punkt im rechten Bild passt.

Es gibt bestimmte Ereignisse, die für die Bestimmung von Disparitäten kritisch sind. Es kann passieren, dass Punkte im linken Bild im rechten Bild durch andere Punkte verdeckt werden, weil beide Kameras unterschiedliche Perspektiven auf die Szene haben. Je nach Blickwinkel existieren Bereiche, die in einem Bild zu erkennen sind und im anderen Bild versteckt liegen. Somit ist keine Disparität angebar.

Weiterhin müssen die Algorithmen zur Disparitätsermittlung darauf achten, dass die Pixelanordnung in beiden Bildern nicht gleich bleibt. Wie in der Abbildung 3.4 kann es geschehen, dass Pixel aufgrund von verschiedenen Blickwinkel anders geordnet sind. Die geordnete Struktur ist aufgelöst. Deswegen kann nicht davon ausgegangen werden, dass sobald ein Korrespondenzpaar an Bildpunkten gefunden worden ist, die Nachbapixel dadurch ebenfalls schon lokalisiert sind.

Verdeutlicht wird diese Problematik am Beispiel von Spiegelungen. Diese variieren von Bild zu Bild und sind bei verschiedenen Blickwinkeln an unterschiedlichen Stelle zu finden. Sie verhindern gegebenenfalls eine Zuordnung von Korrespondenzpunkten, da ein gesuchter Punkt im linken Bild möglicherweise aufgrund einer Spiegelung an diesen Koordinaten im rechten Bild verdeckt wird. Sie erschweren damit eine stabile Korrespondenzanalyse.



Abbildung 3.4: *Problematik bei der Disparitätsbestimmung: In der linken Aufnahme ist das rechte Fenster ganz auf der linken Seite der Hausstrebe. Wohingegen im rechten Bild das rechte Fenster komplett auf der linken Seite der Strebe ist [14].*

In OpenCV sind zur Zeit drei Algorithmen zur Bestimmung der Disparität implementiert. Der Algorithmus, der im Qualität/Performance-Vergleich am Besten abschneidet, soll in diesem Kapitel kurz erläutert werden. Auch der einfachste Algorithmus soll hier kurz zur Verdeutlichung der Funktionsweise und zum besseren Verständnis von Disparität angeschnitten werden.

3.1 Block Matching

Der Block Matching Algorithmus, kurz BM, basiert auf einer Idee von Kurt Konolige [4] und implementiert die Korrespondenzsuche mittels eines SAD-Window, welches auch die benachbarten Bildpunkte des untersuchten Pixels mit einbezieht. SAD steht für Sum of absolute Differences, was bedeutet, dass als Maß für die Ähnlichkeit der Pixelwerte der mittlere quadratische Fehler verwendet wird. Der BM wird in drei Schritte unterteilt:

1. **Prefiltering:** Die Lichthelligkeit beider Bilder wird normalisiert und

Strukturen werden hervorgehoben.

2. **Korrespondenzsuche:** Nimmt man einen Pixel (x_0, y_0) im linken Bild, geht man auf der gleichen Zeile in das rechte Bild und beginnt dort ebenfalls an den Koordinaten $(x_0 + \text{MinDisparity}, y_0)$, wobei *minDisparity* die minimale Verschiebung der Suche in x-Richtung. Dort sucht man nach dem korrespondierenden Punkt. Dabei werden eine vorgegebene Anzahl an Pixeln in negativer x-Richtung, NumDisparities, bis MaxDisparity untersucht und deren mittlerer Fehler bestimmt. Die Differenz zwischen x_0 und $x_0 + \text{MinDisparity} + \text{NumDisparity}$ wird als δ definiert.

$$\begin{aligned}\delta &= x_0 + \text{MinDisparity} + \text{NumDisparity} - x_0 \\ &= \text{MinDisparity} + \text{NumDisparity}\end{aligned}\tag{3.5}$$

Der mittlere quadratische Fehler bestimmt sich dann durch [14]:

$$\begin{aligned}SAD(x, y, \delta) &= \frac{1}{nm} \sum_{i=-k}^k \sum_{j=-k}^k (|I_l(x+i, y+j) - I_r(x+i+\delta, y+j)|^2) \\ &\quad \text{mit } n = m = 2 * k + 1\end{aligned}\tag{3.6}$$

Dargestellt wird die Vorgehensweise des Block Matching Algorithmus in Abbildung 3.5. Überprüft werden alle Pixel, die in einem vorher vorgegebenen Bereich, dem sogenannten Window oder Block, liegen. Für jeden untersuchten Punkt wird für das Ähnlichkeitsmaß der mittlere quadratische Fehler berechnet und in einer Matching-Funktion gespeichert. Der Punkt mit dem höchsten Wert in dieser Funktion wird als gesuchter Punkt deklariert. Wie in der folgenden Abbildung dargestellt, ist die Matching-Funktion nicht diskret. Zwischen den Pixeln wird interpoliert. Wenn man dies für alle Punkte im

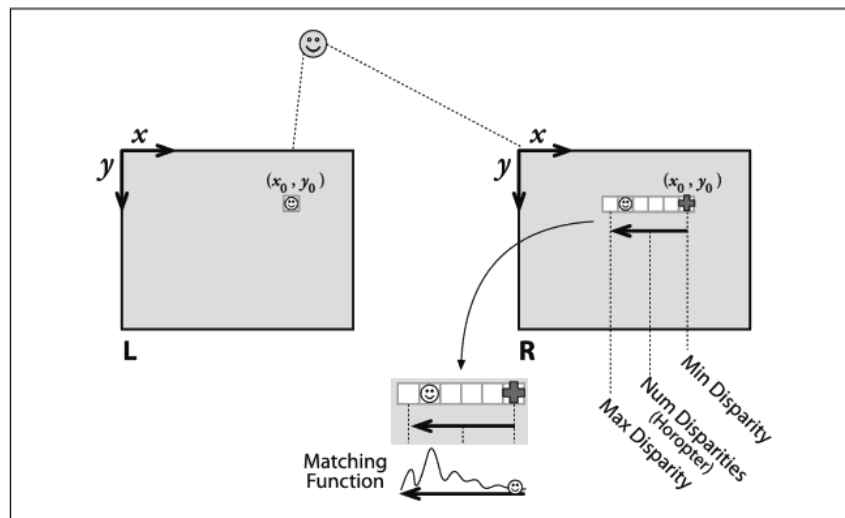


Abbildung 3.5: Funktionsweise des Block Matching Algorithmus: Durch Bilden eines Pixelblocks an den Koordinaten des referenzierten Bildpunktes im rechten Bild wird der Suchbereich für den korrespondierenden Punkt eingegrenzt. Bildpunkte innerhalb des Blockes werden auf Ähnlichkeit geprüft. Der Pixel mit der höchsten Ähnlichkeit ist der korrespondierende Punkt [2].

linken Bild durchführt, erhält man die Disparitätsmatrix. Diese wird mit den zugeordneten Definitionen eines Disparitätbildes in Abbildung 3.7 visualisiert.

Um Ausreißer zu minimieren, werden die mittleren absoluten Fehler eines Blockes mit einem Medianfilter bearbeitet, der verhindert, dass zu große Abweichungen zwischen den einzelnen Disparitäten entstehen [14].

3. **Postfiltering:** Falsche oder unpassende Korrespondenzzuordnungen werden eliminiert. Es wird eine obere Grenze für den SAD definiert. Wenn ein berechneter SAD-Wert diese Toleranzgrenze überschreitet, wird an dieser Stelle keine Disparität bestimmt. Es wird deklariert, dass dort keine Zuordnung von korrespondierenden Punkten stattfinden konnte.

Die Abbildung 3.6 zeigt Aufnahmen von zwei Kameras, die die gleiche Szene aus unterschiedlichen Perspektiven erfasst haben. Zu sehen sind Personen, die sich auf einer Tribüne in Richtung Ausgang bewegen. Die Bilder sind bereits entzerrt und rektifiziert. Der BM ist der simpelste in OpenCV

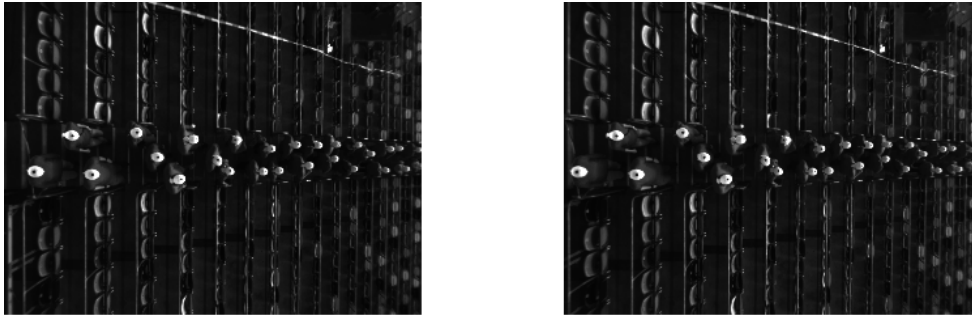


Abbildung 3.6: *Rektifiziertes Stereobildpaar: Auf den Bildern sind Personen dargestellt, die sich auf einer Tribüne Richtung Ausgang bewegen. Die beiden rektifizierten Bilder sind von einer linken und rechten Kamera, die eine Stereokamera bilden, aufgenommen worden und unterscheiden sich augenmerklich kaum.*

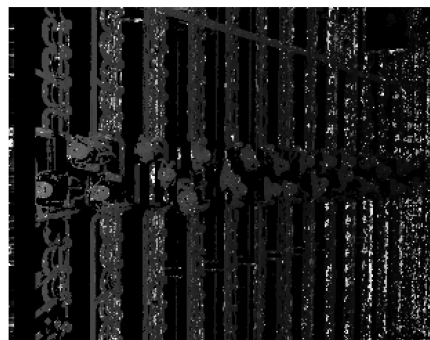


Abbildung 3.7: *Disparitätsbild durch BM-Algorithmus: Das dargestellte Disparitätsbild ist sehr ungenau. Es existieren viele schwarze Bereiche, die eine fehlende Zuordnung von Korrespondenzpunktpaaren repräsentiert.*

implementierte Algorithmus zur Ermittlung von Disparitäten. Er hat eine gute Performance. Allerdings ist das Disparitätsbild sehr ungenau. Korrespondierende Punkte werden nicht exakt und präzise zugeordnet. Viele Punkte können nicht entsprechend gefunden werden, sodass viele Bereiche im Disparitätsbild schwarz erscheinen (siehe Abbildung 3.7), was auf eine fehlende Zuordnung von korrespondierenden Punkten weist. Die entsprechenden Punkte im rechten Bild konnten nicht lokalisiert und somit keine Disparität berechnet werden. Der SAD-Wert der überprüften Punkte ist gegenfalls zu groß und überschreitet die gegebene obere Grenze. Eine angemessene Erweiterung ist allerdings langwierig, wenn das optimale Resultat erzielt werden soll. Es fällt schwer zu entscheiden, welcher Bereich einen weiteren Abstand zur Kamera hat. Das zentrale Problem des BM ist die Bestimmung einer geeigneten Fenstergröße. Um eine zuverlässige Disparitätsbestimmung zu garantieren, ist ein genügend breites Fenster nötig, damit große Intensitätsunterschiede mit einfließen können. Dennoch umfasst ein breites Fenster viele Pixel, die ähnliche Intensitäten haben und somit einen gleichen mittleren absoluten Fehler besitzen, sodass nicht definiert werden kann, welcher der korrespondierende Bildpunkt ist und keine Disparität bestimmbar ist [3].

3.2 Semi-Global Block Matching

Der Semi-Global Block Matching Algorithmus, oder kurz SGBM, stammt von Heiko Hirschmüller [5] und bedient sich der dynamischen Programmierung, das heisst, dass ein komplexes Problem in weniger komplexe Teilprobleme zerlegt wird. Bei diesem Algorithmus wird die Korrespondenzsuche durch die Minimierung einer globalen Matchingkosten-Funktion approximiert, wobei Pixel innerhalb eines Blockes überprüft werden. Die Matchingkosten werden in horizontaler, vertikaler und diagonaler Richtung akkumuliert und mit einem kantenerhaltenden Glattheitskriterium verknüpft [14]. Dieses Kriterium basiert auf der Tatsache, dass sich der Abstand zur Kamera von Nachbapixeln kontinuierlich ändert, sodass sich die Disparität örtlich benachbarter Bildpunkte ebenfalls nur innerhalb ei-

ner Toleranzgrenze ändert [3]. Ist der Tiefensprung zwischen benachbarter Bildpunkte zu groß, wird keine Disparität bestimmt. Auch hier existieren ähnliche Pre- und Postfilterungsmöglichkeiten, um die Resultate zu verbessern.

Der SGBM gliedert sich in vier Stufen [10]:

1. **Kostenermittlung:** Die Matchingkosten-Funktion wird blockweise mittels des Subpixel-Algorithmuses von Birchfield und Tomasi [1] berechnet.

Die Kostenfunktion $C_{BT}(p, d)$ wird für einen Pixelblock p von $minX$ bis $maxX$, der linken und rechten Windowgrenze, berechnet. Dabei ist d die aktuelle Disparität in den rektifizierten Bildern und x repräsentiert den Pixel im aktuellen Block. Die Berechnung der Kostenfunktion wird für alle Disparitäten zwischen dem gegebenen Maximum und Minimum der Disparität, $minD$ und $maxD$, durchgeführt.

$$C_{BT} = (x - minX)maxD - minD + (d - minD) \quad (3.7)$$

2. **Kostenaggregation:** Verbindet alle Kosten der Nachbarpixel und wird in alle Richtungen vom untersuchten Bildpunkt aus berechnet um den Pixel mit dem niedrigsten Disparitätswert zu finden. Berechnet werden die Kosten durch eine globale Energiefunktion.

$$E(D) = \sum_p (C_{BT}(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1]) \quad (3.8)$$

mit $P_1 \leq P_2$

Die Kostenaggregation ist dazu da, um falsche Korrespondenzen zu glätten und die richtigen Korrespondenzen herauszufiltern. Die Energie des Disparitätbildes wird durch die Summe aller Matchingkosten

für eine bestimmte Disparität D berechnet. Der Strafterm für kleine Änderung in der Disparität P_1 ist konstant, wohingegen der Strafterm für eine starke Änderung der Disparität P_2 variiert, um Diskontinuitäten und Intensitätsänderungen abzufangen. Durch Minimierung der Energiefunktion 3.8 ist dann das Matching möglich. Da die 2D-Energieminimierung NP-vollständig ist, wird sie auf ein 1D-Problem zurückgeführt. Dies geschieht durch Aufteilen des zu untersuchenden Blockes in verschiedene Pfade. Abbildung 3.8 zeigt, wie die Energie dann mittels der Pfade in verschiedene Richtungen berechnet wird. Die Pfadkosten L_r eines jeden Pfades berechnen sich

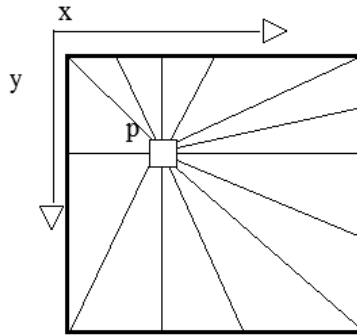


Abbildung 3.8: *Pfadanalyse: Mittels der Aufteilung des zu untersuchenden Pixelblockes in verschiedene Pfade wird die Minimierung der Energiefunktion $x.x$ von einem 2D-Problem auf ein 1D-Problem reduziert und ist damit in polynomialer Zeit berechenbar.*

dann durch die entsprechende Matchingkostenfunktion 3.7 entlang dieses Pfades. Abgezogen werden davon die minimalen Kosten der vorherigen Pixel k , um die Zahlen so klein wie möglich zu halten. Da diese konstant sind, ändert dies nichts an den minimalen Kosten des Pfades.

$$L_r(p, d) = C_{BT}(p, d) + \min(L_r(p - r, d), L_r(p - r, d - 1) + P_1, L_r(p - r, d + 1) + P_1, \min_i L_r(p - r, i) + P_2) - \min_k L_r(p - r, k) \quad (3.9)$$

3. **Disparitätsermittlung:** Berechnet die Disparität auf Grundlage der zuvor ermittelten minimierten Pfadkosten jedes Pixels.
4. **Disparitätsverfeinerung:** Ähnlich der Postfilterung beim BM-Algorithmus.

Der SGBM ist trotz seiner sehr guten Qualität sehr schnell. Er liefert sehr genaue Ergebnisse. Auch die Fehlzuordnung oder fehlende Zuordnung tritt vermindert auf. Nur sehr wenige Komponenten sind schwarz und symbolisieren damit fehlende Korrespondenzen.

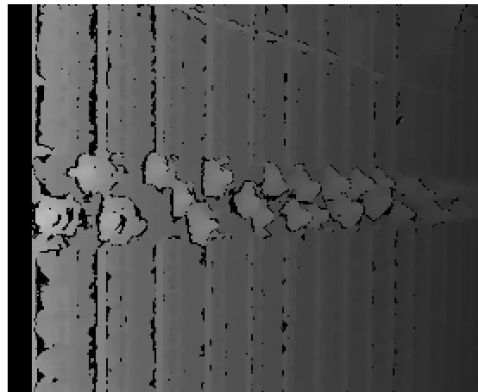


Abbildung 3.9: *Disparitätsbild durch den SGBM-Algorithmus: Weniger Bereiche des Disparitätsbildes sind schwarz. Eine fast flächendeckende Korrespondenzzuordnung wird erwirkt.*

Kapitel 4

Verbund von Stereokameras

Sollen Personen über einen größeren Bereich verfolgt werden oder ist die Höhe der Kameraaufhängung beschränkt, so ist es wünschenswert das beobachtete Gebiet über das Sichtfeld einer Kamera beziehungsweise einer Stereokamera zu erweitern. Aus dieser Motivation soll der Beobachtungsbereich einer Kamera auf weitere Sichtbereiche ausgedehnt werden. Das vorgestellte System beschreibt dabei zur Demonstration eine Minimalkonfiguration eines Stereogrids mit einer Stereokamera und einer den Sichtbereich erweiterten dritten Kamera. Dafür müssen sich die Sichtbereiche der drei monokularen Kameras wie in den bisherigen Vorraussetzungen der vorangestellten Kapiteln überschneiden. In Abbildung 4.1 ist der grundsätzliche Kameraaufbau zu erkennen.

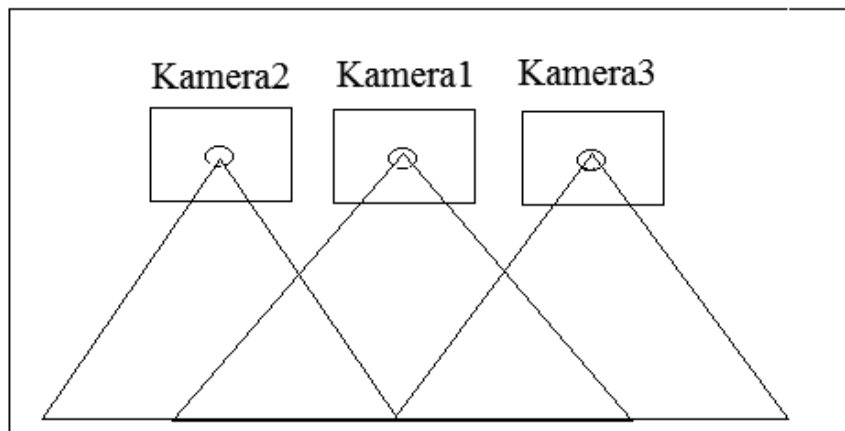


Abbildung 4.1: *Kameragridausrichtung: Die drei monokularen Kameras werden nebeneinander angebracht. Von jeweils zwei Kameras überschneidet sich der Sichtbereich, wobei die mittlere Kamera zwei überlappende Sichtbereiche hat und sowohl einen Teil der linken als auch einen Teil der rechten Beobachtungsfläche sieht.*

Wichtig ist nun jedoch, dass der Überdeckungsbereich so gering wie möglich ist. Dabei muss das Überlappungsgebiet groß genug sein, um eine Korrespondenzsuche zu gewährleisten. Jeder Raumpunkt muss von mindestens zwei Kameras zu sehen sein. Allerdings muss der überschneidende Bereich auch so klein wie möglich sein, um abweichende Bilder und damit eine erweiterte Beobachtungsfläche zu erzielen.

4.1 Aufbau

Mittels eines minimalen Grids wird in diesem Kapitel die Thematik der Vergrößerung der Beobachtungsfläche behandelt. Zur Verfügung stehen drei monokulare Kameras. Sie stellen das minimale Grid dar. Die drei Kameras werden in horizontaler Linie angebracht vergleiche Abbildung 4.2. Sie liegen somit nahezu auf gleicher Höhe und unterscheiden sich in ihrer x-Position und Rotation zueinander. Zur Konstruktion des Stereogrids



Abbildung 4.2: Kameraanbringung: Die drei monokularen Kameras werden nebeneinander auf nahezu gleicher horizontaler Linie angebracht. Die linke und mittlere Kamera gehören zum gleichen Modell. Die rechte Kamera ist ein neueres Modell desselben Herstellers.

ist keine horizontal gleiche Ausrichtung der Kameras notwendig. Sie können auch versetzt und in unterschiedlichen x- und y-Positionen angebracht sein. Lediglich aus Einfachheit und zur Übersichtlichkeit werden die drei Kameras auf gleicher Höhe angebracht. Die Ergebnisse und Konstruktionsschritte sind dennoch auf alle Kamerazusammenschlüsse übertragbar, solange sich die Sichtbereiche, wie in den Voraussetzungen geschildert, überschneiden.

4.2 Konstruktion

Bei der Konstruktion des Stereogrids müssen zunächst die gleichen Arbeitsschritte und Prozesse wie bei zur Stereokalibrierung durchgeführt werden. Viele Eigenschaften und geometrischen Beziehungen der drei Kameras zueinander und dadurch auch des Stereogrids lassen sich in ähnlicher Weise auf die bisherigen Modelle und Algorithmen der Stereokalibrierung beziehen und müssen lediglich an einigen Stellen angepasst und erwei-

tert werden. Bei der Konstruktion des Stereogrids werden folgende zwei Schritte durchgeführt:

1. **Entzerrung:** Auch bei der Konstruktion des Stereogrids ist es notwendig vorhandene Verzerrungen der drei Kameras zu entfernen, um die Weiterverwendung und Analyse der Kameraaufnahmen zu vereinfachen. Die extrinsischen und intrinsischen Parameter jeder Kamera werden wie in Kapitel 2.1 berechnet. In diesem Schritt muss noch nicht auf das Grid eingegangen werden, da die Entzerrung ein unabhängiger Prozess ist, der für jede Kamera einzeln durchzuführen ist. Er ist entbunden von jeglichen geometrischen Beziehungen und Abhängigkeiten zwischen den Kameras.

Wie in Kapitel 2.1 geschildert, wird hier zur Berechnung der gesuchten Parameter das Schachbrett als Ablichtungsobjekt benutzt. Dieses liefert genügend Objekt- und Bildpunkte und garantiert somit eine Lösung des nichtlinearen Gleichungssystems 2.3 und dadurch die Approximation der nötigen Kameraparameter.

2. **Ausrichtung und perspektivische Transformation:** Nach Berechnung der extrinsischen und intrinsischen Parameter jeder Kamera müssen die Geometrien und die Ausrichtung der Kameras zueinander ermittelt werden. Dazu werden aus den drei Kamera jeweils Paare aus zwei Kameras gebildet werden, um die Eigenschaften und Transformationsvorschriften der Stereorekonstruktion anwenden zu können. Die linke Kamera und die mittlere Kamera (im Folgenden: Paar 1), sowie die mittlere und die rechte Kamera (im Folgenden: Paar 2) werden zu einem Kamerapaar wie in Abbildung zusammengefasst 4.3.

In den sich überschneidenden Sichtbereiche wird ein Schwarz/Weiß-Schachbrett gelegt, auf das beide Kameras eines Paares Sicht haben. Dieses Schachbrett ermöglicht, die Berechnung von Orientierung und Position der beiden Kameras zueinander. Wie in Kapitel 2.1 beschrieben, können durch die erhaltenen Bild- und Objektpunkte des Schachbrettes beider Kameras die geometrischen Zusammenhänge und Abhängigkeiten ermittelt werden, das heißt die perspektivische

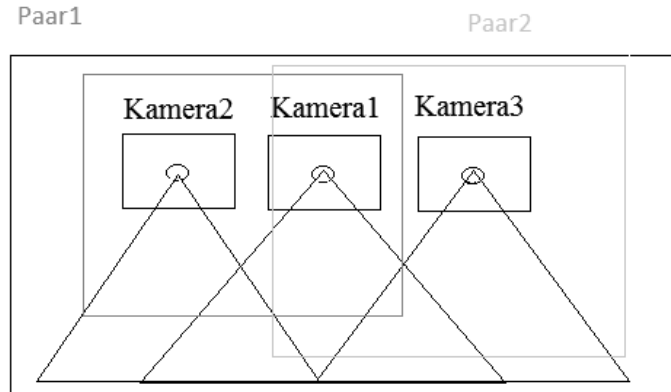


Abbildung 4.3: *Bilden von Kamerapaaren: Jeweils zwei Kameras werden zu einem Paar zusammengefasst. Dabei ist die mittlere Kamera (Kamera1) in beiden Kamerapaaren enthalten.*

Projektion des einen Kamerakoordinatensystems in das andere Koordinatensystem, sodass beide Kameras das gleiche Koordinatensystem verwenden. Die Projektion des einen in das andere Koordinatensystems setzt sich aus den extrinsischen Parametern der beiden Kameras zusammen, was im Wesentlichen durch eine Rotation und Translation wie in Gleichung 4.2 geschieht.

$$P_m = R_{lm}P_l - T_{lm} \quad (4.1)$$

P_l symbolisiert die Bildpunkte der linken Kamera, welche durch Transformation in das Koordinatensystem der mittleren Kamera gelegt werden. P_m stellt somit die Punkte P_l bezüglich des ersten Koordinatensystems dar und beschreibt diese in Koordinaten des mittleren Koordinatensystems. Die Rotation R_{lm} und die Translation T_{lm} definieren die Projektion des linken Kamerakoordinatensystems in das mittlere. Für das Stereosystem 2, welches aus der mittleren und rechten Kamera besteht, wird ebenfalls die Projektion des rechten in das mittlere Kamerakoordinatensystems berechnet.

$$P_m = R_{rm}P_r - T_{rm} \quad (4.2)$$

Die Ausrichtung und Orientierung der linken und rechten Kamera bezüglich der mittleren Kamera können somit ermittelt werden. Das mittlere Kamerakoordinatensystem bildet das Referenzkamerakoordinatensystem. Die drei Kameras haben somit alle das gleiche Koordinatensystem, welches durch das mittlere Kamerakoordinatensystem repräsentiert wird.

In der Abbildung 4.4 wird die Konstruktion des Stereogrids an einem Beispiel erläutert. Die Aufnahmen und Observierungen einer Szene durch das selbst konstruierte Grid zeigen, dass sich die Sichtbereiche der Kameras überschneiden und manche Gebiete der Szene in mehreren Kamerabildern erkennbar sind. Andere Gebiete sind dagegen nur in der Beobachtungsfläche einer Kamera zu sehen. Die Aufnahme müssen nun so verarbeitet werden, dass sie eine vergrößerte Beobachtungsfläche bieten und mehrfach zu sehende Bereiche ausgeblendet und nur einmal dargestellt werden. Die Bilder der drei Kameras sollen “zusammenschmelzen”, und nach der Ausrichtung der Kameras so erscheinen, als wären sie von einer einzelnen Kamera aufgenommen. Ein kontinuierliches Bild entsteht.

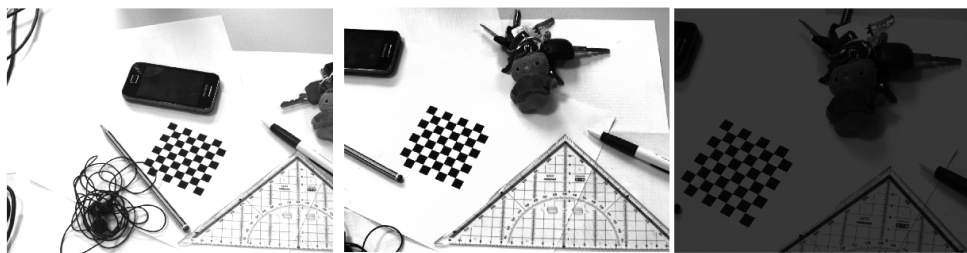


Abbildung 4.4: Gridaufnahme: Die Bilder zeigen eine aufgenommene Szene durch das Grid und somit durch drei monokulare Kameras. Man sieht, dass bestimmte Bereiche in mehreren Kameras zu sehen sind. Andere Teile sind nur innerhalb eines Kamerasichtbereiches sichtbar.

Kapitel 5

Zusammenfassung und Ausblick

Zur Bearbeitung der Thematik zur Stereokalibrierung stehen zwei identische monokulare Kameras zu Verfügung, die ein Stereosystem bilden. Aufgrund der Objektive und ihrer Eigenschaften enthalten die aufgenommenen Kamerabilder Verzerrungseffekte wie zum Beispiel kissenförmige Verzeichnungen, die entfernt werden müssen, um eine einfachere Weiterverarbeitung basierend auf dem Modell einer Lochkamera zu ermöglichen. Bei der Entzerrung beruft man sich auf das Modell der Lochkamera, welche aufgrund der fehlenden Linse unverzeichnete Aufnahmen macht. Mit Hilfe dieses Modells können die Verzerrungen weitestgehend extrahiert werden. Zwar sind noch Verzerrungseffekte nach der Entzerrung vorhanden, dennoch beeinflussen diese die weiteren Kalibrierungsschritte und die Disparitätsbestimmung nur minimal. Die Entzerrung bestimmt die Kalibrierungsparameter einer einzelnen Kamera und ist somit vom Stereosystem und dem Prozess der Stereokalibrierung unabhängig. Sie ist allerdings für die Stereokalibrierung als Vorverarbeitungsschritt notwendig. Die Epipolargeometrie beschreibt die geometrischen Beziehungen zweier Lochkameramodelle, die aus unterschiedlicher Perspektive das gleiche Objekt abbilden. Durch die Epipolargeometrie können die äußeren Orientierungen der Kameras zueinander bestimmt und somit ein einheitliches

Koordinatensystem beider Kameras definiert werden. Dabei setzt man das Koordinatensystem der linken Kamera als das Referenzkoordinatensystem und transformiert das zweite hinein. Durch die Anpassung an ein gemeinsames Koordinatensystem kann die parallele Ausrichtung beider Kameras zueinander gewährleistet werden. Nach der Rektifizierung stehen beide Kameras parallel zueinander. Der Suchbereich für die Korrespondenzanalyse ist auf eine Zeile eingeschränkt und wurde von einem 2D auf ein 1D-Problem reduziert. Der Abstand eines Bildpunktes zur Kamera wurde über den Zusammenhang der Disparitäten bestimmt. Dabei wurden bei der Disparitätsermittlung zwei Algorithmen miteinander verglichen. Der Semi Global Block Matching Algorithmus hat sich dabei dem Block Matching Algorithmus gegenüber trotz ähnlich guter Effizienz als präziser herausgestellt. Dieser kontrolliert Punkte im rechten Bild nicht nur in x-Richtung zur Findung des Korrespondenzpunktes, sondern durchsucht verschiedene Pfade zum Beispiel die Diagonalen, um gegebene Bedingungen zu erfüllen und in die Suche miteinzubinden.

Das zu konstruierende Stereogrid besteht aus drei monokularen Kameras. Dabei gehören zwei zu dem bereits erläuterten Stereosystem. Das Modell ist identisch. Die dritte Kamera ist zwar vom gleichen Hersteller, ist aber ein neueres Modell und besitzt ebenfalls ein anderes Objektiv. Die Verzerrungseffekte, die durch die Linse verursacht werden, und die innere Geometrie der Kamera unterscheidet sich somit.

Bei der Konstruktion eines Stereogrids sind zunächst die gleichen Kalibrierungsschritte in abgewandelter Weise wie bei der Stereokalibrierung notwendig. Der Entzerrungsschritt bleibt erhalten, da dies unabhängig vom Stereogrid und der Orientierungen der Kameras zueinander geschieht. Er ist speziell auf eine Kamera zugeschnitten und berechnet für diese neben den Verzerrungskoeffizienten auch die Kameramatrix sowie die extrinsischen Parameter. Anschließend werden die drei Kameras in zwei Stereopaare zusammengeschlossen, um die vorher erlangten Erkenntnisse und vorgestellten Algorithmen zur Stereokalibrierung anwenden zu können. Das Koordinatensystem der mittleren Kamera repräsentiert das Referenzkoordinatensystem in das die anderen beide transformiert werden. Die Kamerakoordinatensysteme der rechten und linken Kamera werden über die Ausrichtung relativ zur mittleren Kamera in das Referenzkoordinatensys-

tem projiziert. Das mittlere Kamerakoordinatensystem bietet eine einheitliche Koordinatenrelation der verschiedenen Kamerakoordinaten. Die Kameras haben somit ein gemeinsames Weltkoordinatensystem. Nach der Transformation aller Bildpunkte der Kameras in dasselbe Kamerakoordinatensystem ist eine vergrößerte Beobachtungsfläche darstellbar.

5.1 Ausblick

Bei den vorgestellten Modellen und Berechnungen wurde bisher keinerlei Rücksicht auf Approximationsfehler und Fehlerentwicklung genommen. Es wurde nicht berücksichtigt, wie sich ein erster Approximationsfehler auf die weiteren Berechnungen im weiteren Prozess verhält und inwiefern er auf das Ergebnis Einfluss hat. Mittels einer Fehleranalyse könnte untersucht werden, wie stark welche Fehler die Resultate beeinflussen und eine Verbesserung möglich wäre. Neben der Untersuchung von Fehlerarten muss ebenfalls überprüft werden, wie sich diese Fehler fortpflanzen. Um dann eine Verbesserung zu unterstützen, muss analysiert werden, wie sich die Fehler minimieren und somit das ganze System beeinflussen lassen.

Eine weitere Fortsetzung der Arbeit wäre eine vollständige 3D-Rekonstruktion der gewonnenen Disparitätswerte, um letztendlich eine Personenerkennung realisierbar zu machen. Dafür sind allerdings neben weiteren Modellen vor allem Definitionen und Eigenschaften für Personengrößen und -formen etc. aufzustellen. Es ist zu modellieren, wie eine Person “aussieht“, durch welche Formen sie sich approximieren lässt und welche Merkmale sie von anderen Objekten unterscheidet. Erst anhand dieser Definitionen kann dann aus den 3D-Punkten erkannt werden, ob das detektierte Objekt eine Person ist oder zum Beispiel ein Schrank.

Anhang A

Implementierung

Die Implementierung der Arbeit erfolgt in C++ mit Hilfe der Bibliothek QT Version 4.7.2 zur plattformübergreifenden Programmierung der graphischen Oberfläche [6]. Desweiteren wurde die Opensource-Bibliothek OpenCV Version 2.2 verwendet, die für das Bildverstehen (Computer Vision) entwickelt wurde und schon viele Algorithmen und Funktionen zur Stereokalibrierung sowie Bildverarbeitung zur Verfügung stellt [2]. Das Programm arbeitet mit USB-Kameras vom Hersteller IDS GmbH [17]. Die Implementierung dieser Arbeit gliedert sich in zwei Programme. Ein Programm beinhaltet die Stereorekonstruktion und die anschließende Disparitätsermittlung. Das zweite Programm behandelt die Konstruktion des Verbundes von Stereokameras.

Stereorekonstruktion

Die Abbildung A.1 zeigt den Ablauf der Stereorekonstruktion mit anschließender Disparitätsbestimmung in einem Blockdiagramm. Zur Stereorekonstruktion sind zwei monokulare Kameras so aufzubauen, dass sich ihre Sichtbereiche größtenteils überschneiden und sie ein Stereosystem bilden. Als Kalibrierungsobjekt dient wie in Kapitel 2 beschrieben ein Schwarz/Weiß-

ANHANG A. IMPLEMENTIERUNG

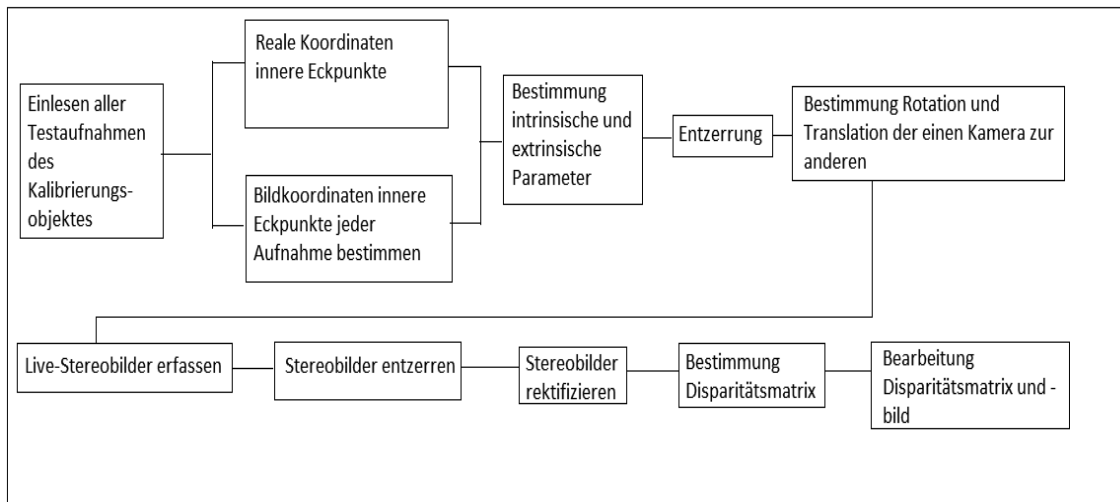


Abbildung A.1: *Stereorekonstruktion: Das Blockdiagramm zeigt den Ablauf der Stereorekonstruktion mit anschließender Disparitätsbestimmung.*

Schachbrett. Dabei ist darauf zu achten, dass das Schachbrett mit einem weißen Rand umrahmt ist, da die inneren Ecken der schwarzen Randquadrate ansonsten nicht gefunden werden können. Von dem Schachbrett sind mehrere Testaufnahmen aus verschiedenen Perspektiven zu erfassen. Das Schachbrett muss in den jeweiligen Aufnahmen komplett sichtbar sein, damit die Korrespondenzsuche später realisiert werden kann. Die realen Koordinaten der inneren Ecken des Schachbrettes werden ermittelt, indem die Breite eines Quadrates ausgemessen wird und die inneren Ecken in der Breite und in der Höhe gezählt werden. Sie bestimmen die Objektpunkte des Schachbrettes. Zusätzlich werden in jeder Testaufnahme die inneren Ecken der Schachbrettquadrate gesucht. Diese definieren die Bildpunkte des Schachbrettes. Durch Kenntnis der Bild- und Objektpunkte des Schachbrettes werden die intrinsischen und extrinsischen Parameter jeder Kamera berechnet, wodurch auch die Verzerrungskoeffizienten ermittelt und somit die Entzerrung der Kameraaufnahmen durchgeführt werden kann. Nach der Entzerrung aller Testaufnahmen werden die Orientierungen der Kameras zueinander über die Epipolargeometrie bestimmt. Die Stereobilder werden rektifiziert, sodass sich Bildzeilen in den jeweiligen Stereo-

aufnahmen entsprechen. Nach der Rektifizierung sind alle nötigen Parameter vorhanden, um die Disparitätsermittlung auszuführen. Diese wird an einem Live-Stereobild durchgeführt. Dazu wird von jeder der beiden Kameras ein Live-Bild der gleichen beobachteten Szene erfasst, welches mittels der zuvor berechneten Kalibrierungsparameter entzerrt und rektifiziert wird. Da sich der Kameraaufbau während der Erfassung der Testaufnahmen nicht ändert, müssen auch die berechneten Parameter und Transformationen nicht angepasst werden. Über den Semi-Global Block Matching Algorithmus wird anschließend die Disparität eines jeden Pixels, soweit bestimmbar, ermittelt und in einer Disparitätsmatrix gespeichert. Diese wird in einem Disparitätsbild dargestellt. Neben dem Disparitätsbild werden in einer graphischen Oberfläche verschiedene Slider zur Verfügung gestellt, um eine Veränderung und Variabilität der Parameter zur Disparitätsbestimmung durch den Benutzer zu ermöglichen. Neben der Variabilität der Blockgröße oder der minimalen Disparität ist auch eine Einstellung der Toleranzgrenze von Disparitätsunterschieden benachbarter Pixel möglich. Die Disparitätsmatrix und somit das Disparitätsbild sind dadurch vom Benutzer flexibel veränderbar, um den Einfluss der unterschiedlichen Parameter auf das Resultat zu verdeutlichen und um das Resultat gegebenenfalls individuell anzupassen.

Stereogridkonstruktion

Bei der Konstruktion des Verbundes von Stereokameras werden die drei monokularen Kameras so ausgerichtet, dass sich ihre Sichtbereiche geeignet überschneiden. Jeder Objektpunkt muss in mindestens zwei der Kamerasichtbereiche liegen. Dennoch darf die Überschneidung der Beobachtungsflächen nicht zu groß sein, um eine Vergrößerung dieser zu gewährleisten.

Die Abbildung A.2 stellt den Prozess der Stereogridkonstruktion in einem Blockdiagramm dar. Als Kalibrierungsobjekt wird ebenfalls das Schwarz/Weiß-Schachbrett verwendet. Dieses wird in die überlappenden Sichtbereiche gelegt und ist damit jeweils in mindestens zwei Sichtbereiche zu erken-

ANHANG A. IMPLEMENTIERUNG

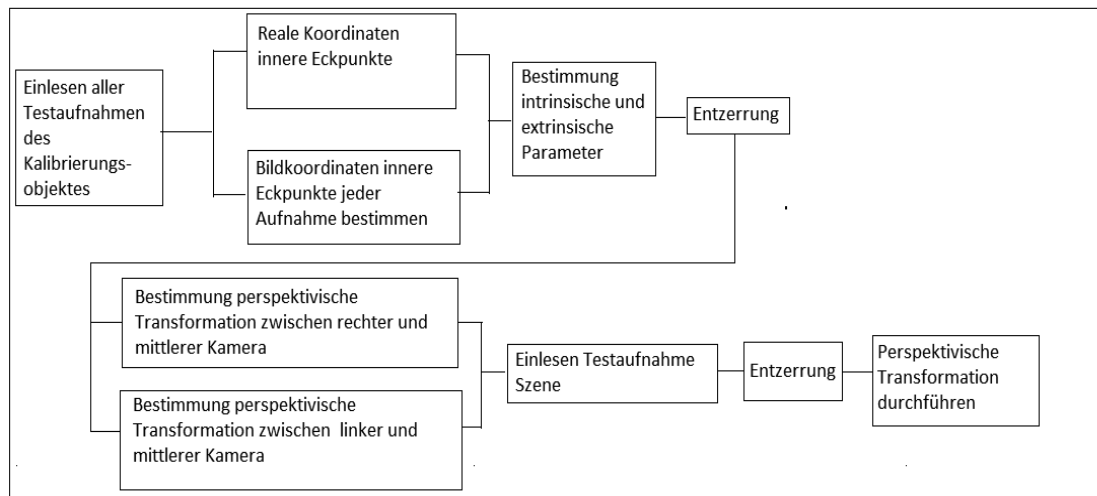


Abbildung A.2: Stereogridkonstruktion: Das Blockdiagramm zeigt den Ablauf der Konstruktion eines Verbundes von Stereokameras.

nen. Die Bestimmung der Objekt- und Bildpunkte des Schachbrettes, sowie die anschließende Ermittlung aller nötigen Kameraparameter (intrinsische, extrinsische Parameter und Verzerrungskoeffizienten) erfolgt wie in Kapitel A.1 erläutert. Die Testaufnahmen werden anhand dieser Parameter entzerrt.

Um die Transformation der drei Kamerakoordinatensysteme in ein einheitliches Koordinatensystem werden jeweils zwei Kameras zu einem Paar zusammengefasst und zueinander ausgerichtet. Die mittlere Kamera ist dabei in jedem Paar vorhanden, sodass in dem jeweiligen Kamerapaar die perspektivische Transformation in das mittlere Kamerakoordinatensystem erfolgt. Die Rotationsmatrizen und Translationsvektoren der rechten und linken Kameras bezüglich der mittleren Kamera werden berechnet. Danach werden die drei Aufnahmen einer vom Stereogrid observierten Szene eingelesen. Da alle relevanten Parameter und Transformationsmatrizen bekannt sind, können die Aufnahmen entzerrt und entsprechend ausgerichtet werden.

Anhang B

Tabellen

Tabelle B.1: Eigenschaften der monokularen Kameras der Firma IDS [17]

| Eigenschaften | <i>UI – 2230SE</i> | <i>UI – 1240SE</i> |
|-----------------------|--------------------|--------------------|
| Interface | USB | USB |
| Objektivanschluss | C-Mount | C-Mount |
| Sensor Technologie | CCD (Sony) | CMOS (e2v) |
| Auflösung (h x v) | 1024 x 768 | 1280 x 1024 |
| Farbtiefe | 8bit (12bit ADC) | 8bit (10bit ADC) |
| Auflösung/Pixelklasse | XGA | 1.3 Megapixel |
| Optische Fläche | 4.762 x 3.571mm | 6.784 x 5.427mm |
| Abmessungen HxWxL | 34 x 32 x 34.40mm | 34 x 32 x 27.40mm |
| Gewicht | 74.00g | 62.00g |

ANHANG B. TABELLEN

Literaturverzeichnis

- [1] S. Birchfield and C. Tomasi, *Depth Discontinuities by pixel-to-pixel stereo*, International Journal of Computer Vision. Second Edition, 1994.
- [2] G. Bradski and A. Kaehler, *Learning OpenCV. Computer Vision with the OpenCV Library*, O'Reilly Media. First Edition, 2008.
- [3] O. Schreer, *Stereoanalyse und Bildsynthese*, Springer-Verlag, 2005
- [4] K. Konolige und M. Agrarwal, *Frame-Frame Matching for Realtime Consistent Visual Mapping*
- [5] H. Hirschmueller, *Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information*, IEEE Conference on Computer Vision and Pattern Recognition, 2005
- [6] J. Wolf, *Qt 4.6 - GUI-Entwicklung mit C++: Das umfassende Handbuch*, Galileo Computing, Zweite Auflage, 2010
- [7] Z. Zhang *A flexible new technique for camera calibration*, IEEE Transactions on Pattern Analysis and Machine Intelligence 22: 1330-1334, 2000
- [8] R. I. Hartley, *Theory and practice of projective rectification*, International Journal of Computer Vision 35:115-127, 1998

- [9] R. Y. Tsai, *A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses*, IEEE Journal of Robotics and Automatics3: 323-344, 1987
- [10] S. Dröppelmann, M. Hueting, S. Latour und M. van der Veen, *Stereo Vision using the OpenCV library*, 2010
- [11] A. Kläser, *Kamerakalibrierung und Stereo Vision*, 2005
- [12] S. Graf, *Kamerakalibrierung mit radialer Verzeichnung. Die radiale essentielle Matrix*, Universität Passau, 2007
- [13] R. Hartley und A. Zisserman, *Multiple View Geometry in Computer Vision*, Second Edition, Cambridge University Press, 2004
- [14] G. Föll, *Stereo Vision: Vergleich verschiedener Algorithmen zur Lösung des Korrespondenzproblems*, Hochschule für Angewandte Wissenschaften Hamburg, 20010
- [15] C. B. Lee, *Radial undistortion and calibration on an image array*, Massachusetts Institute of Technology, 2000
- [16] F. Greafe, *Stereoskopische Ladeflächenerkennung und Füllstandsmessung am Beispiel von Häckselgut*, Köln, 2007
- [17] *Industriekameras für die digitale Bildverarbeitung und Visualisierung*, 25. Juli 2011
<http://www.ids-imaging.de>
- [18] *Lochkameramodell*, 25. Juli 2011
<http://www.chemgapedia.de/vsengine/media/vsc/de/ph/14/ep/einfuehrung/geooptik/bilder/lochkamera.png>,
- [19] *OpenCV 2.2 C++ Reference*, 25. Juli 2011
http://opencv.jp/opencv-2.2_org/cpp
- [20] *Wikipedia: Epipolargeometrie*, 25. Juli 2011
<http://de.wikipedia.org/wiki/Epipolargeometrie>

- [21] *Wikipedia: Rektifizierung*, 25. Juli 2011
http://en.wikipedia.org/wiki/Image_rectification
- [22] *Wikipedia: Projektionsmatrix*, 25. Juli 2011
http://de.wikipedia.org/wiki/Projektionsmatrix#Radiale_Verzeichnung

Jül-4341
November 2011
ISSN 0944-2952